length of match

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:24:38 ; Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: US-09-396-196F-1
Perfect score: 25
Sequence: 1 gacattgtcgcaagtcacagaatta 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 199986

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		25	100.0	1041	6	AR029499	AR029499 Sequence
2		25	100.0	1041	6	AR034916	AR034916 Sequence
3		25	100.0	1084	6	A11530	A11530 BioB gene o
4		25	100.0	1121	6	E00893	E00893 Genomic DNA
5		25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
6		25	100.0	5872	6	A38246	A38246 Sequence 1
7		25	100.0	5872	6	A38251	A38251 Sequence 6
8		25	100.0	5872	6	A93674	A93674 Sequence 1
9		25	100.0	5872	6	A93679	A93679 Sequence 6
10		25	100.0	5872	6	AR101809	AR101809 Sequence
11		25	100.0	5872	6	AR101810	AR101810 Sequence
12		25	100.0	11022	1	AE000180	AE000180 Escherich
13		25	100.0	13501	1	AE005258	AE005258 Escherich
14		25	100.0	297816	1	AP002553	AP002553 Escherich
15		21	84.0	5526	1	AF250776	AF250776 Unculture
16		18	72.0	8227	1	AF248314	AF248314 Unculture
17	c	16	64.0	32767	2	AC087305	AC087305 Homo sapi
18		16	64.0	155172	2	AC021110	AC021110 Homo sapi
19	c	16	64.0	156184	2	AC041018	AC041018 Homo sapi
20	c	16	64.0	174815	2	AP001390	AP001390 Homo sapi
21	c	16	64.0	178056	2	AC091589	AC091589 Homo sapi
22	c	16	64.0	196550	9	AC020734	AC020734 Homo sapi
23		15	60.0	561	6	AX136747	AX136747 Sequence
24		15	60.0	1771	6	AR080743	AR080743 Sequence
25		15	60.0	1771	10	MUSINTESUB	L20788 Mus musculu
26	c	15	60.0	1829	11	G07181	G07181 human STS W
27		15	60.0	1949	10	MMINTEGR01	U34627 Mus musculu
28		15	60.0	2340	1	ECHK31IM	X82231 E.coli HK31
29		15	60.0	3140	9	Y029487S18	AY029504 Homo sapi
30	c	15	60.0	3520	6	AX136409	AX136409 Sequence
31	c	15	60.0	3748	6	AX098224	AX098224 Sequence
32	c	15	60.0	4602	9	D13644	D13644 Homo sapien
33		15	60.0	4636	1	AF039582	AF039582 Enterobac
34	c	15	60.0	5218	9	AF152313	AF152313 Homo sapi
35	c	15	60.0	5221	9	AF152315	AF152315 Homo sapi
36	c	15	60.0	5233	9	AF152308	AF152308 Homo sapi
37	c	15	60.0	5251	9	AF152312	AF152312 Homo sapi
38	c	15	60.0	5254	9	AF152306	AF152306 Homo sapi
39	c	15	60.0	5254	9	AF152310	AF152310 Homo sapi
40	c	15	60.0	5257	9	AF152307	AF152307 Homo sapi
41	c	15	60.0	5260	9	AF152305	AF152305 Homo sapi
42	c	15	60.0	5260	9	AF152309	AF152309 Homo sapi
43	c	15	60.0	5260	9	AF152311	AF152311 Homo sapi
44	c	15	60.0	5260	9	AF152314	AF152314 Homo sapi
45	c	15	60.0	5260	9	AF152316	AF152316 Homo sapi
46	c	15	60.0	5260	9	AF152317	AF152317 Homo sapi
47	c	15	60.0	5299	9	AF152303	AF152303 Homo sapi
48	c	15	60.0	5431	9	AF152304	AF152304 Homo sapi
49		15	60.0	5562	9	AB011138	AB011138 Homo sapi
50	c	15	60.0	6567	6	I09636	I09636 Sequence 16
51		15	60.0	6767	9	AB051358	AB051358 Homo sapi
52	c	15	60.0	7053	3	ETE306453	AJ306453 Eimeria t
53	c	15	60.0	11980	1	AE005996	AE005996 Caulobact
54		15	60.0	44778	8	SPBC4F6	AL031534 S.pombe c
55	c	15	60.0	59061	2	AC090749	AC090749 Homo sapi
56		15	60.0	77396	2	AC021779	AC021779 Homo sapi
57		15	60.0	83966	2	AC068634_3	Continuation (4 of
58	c	15	60.0	100000	9	AP000067	AP000067 Homo sapi
59	c	15	60.0	100000	9	AP000132	AP000132 Homo sapi
60	c	15	60.0	100000	9	AP000210	AP000210 Homo sapi
61		15	60.0	101829	9	AP000479	AP000479 Homo sapi
62		15	60.0	103244	3	AC090524	AC090524 Caenorhab
63		15	60.0	103908	2	AC084847	AC084847 Homo sapi
64		15	60.0	104729	9	HS118B18	AL034344 Human DNA
65	c	15	60.0	104921	2	AP000439	AP000439 Homo sapi
66		15	60.0	107397	2	AP000612	AP000612 Homo sapi
67		15	60.0	109831	2	AL139285	AL139285 Homo sapi
68	c	15	60.0	110000	2	AC044784_1	Continuation (2 of
69	c	15	60.0	110000	2	AC068782_2	Continuation (3 of
70		15	60.0	110000	2	AC079846_1	Continuation (2 of

c	71	15	60.0	122615	2	AP003965
c	72	15	60.0	123887	9	AL390739
	73	15	60.0	128495	9	AL356913
	74	15	60.0	131541	2	AC006594
	75	15	60.0	137169	2	AC015781
	76	15	60.0	138033	2	AP002339
c	77	15	60.0	143800	2	AC011875
c	78	15	60.0	143840	2	AC091542
c	79	15	60.0	145942	2	AC080001
	80	15	60.0	146002	9	CNS01RGD
c	81	15	60.0	146847	9	AP000248
c	82	15	60.0	147172	9	AC008688
c	83	15	60.0	147706	8	AC083945
	84	15	60.0	148189	9	AC018796
	85	15	60.0	148582	9	AP001574
	86	15	60.0	148845	9	HS1114A1
	87	15	60.0	150613	2	AP003878
	88	15	60.0	150893	2	AC044871
c	89	15	60.0	152719	2	AC012545
	90	15	60.0	154244	2	AC073463
	91	15	60.0	154547	2	AC016112
c	92	15	60.0	155627	2	AC013608
	93	15	60.0	155822	9	CNS01DUV
c	94	15	60.0	156288	2	AC025647
c	95	15	60.0	158621	2	AC011789
c	96	15	60.0	159866	9	AC016598
	97	15	60.0	160855	2	AC011579
c	98	15	60.0	161474	2	AC016198
	99	15	60.0	162383	9	AC074045
c	100	15	60.0	164113	2	AL391820

AP003965	Homo sapi
AL390739	Human DNA
AL356913	Human DNA
AC006594	Homo sapi
AC015781	Homo sapi
AP002339	Homo sapi
AC011875	Homo sapi
AC091542	Felis cat
AC080001	Homo sapi
AL157915	Human chr
AP000248	Homo sapi
AC008688	Homo sapi
AC083945	Oryza sat
AC018796	Homo sapi
AP001574	Homo sapi
AL035684	Human DNA
AP003878	Oryza sat
AC044871	Homo sapi
AC012545	Homo sapi
AC073463	Homo sapi
AC016112	Homo sapi
AC013608	Homo sapi
AL133370	Human chr
AC025647	Homo sapi
AC011789	Homo sapi
AC016598	Homo sapi
AC011579	Homo sapi
AC016198	Homo sapi
AC074045	Homo sapi
AL391820	Homo sapi

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db     21 GACATTGTCGCAAGTCACAGAATTA 45


RESULT   2
AR034916
LOCUS       AR034916      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db     21 GACATTGTCGCAAGTCACAGAATTA 45


RESULT   3
A11530
LOCUS       A11530        1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db     44 GACATTGTCGCAAGTCACAGAATTA 68


RESULT   4
E00893
LOCUS       E00893        1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
```

```
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db     62 GACATTGTCGCAAGTCACAGAATTA 86


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
```

```
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match             100.0%;  Score 25;  DB 1;  Length 5793;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db   2032 GACATTGTCGCAAGTCACAGAATTA 2056


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
```

```
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a    1552 c    1695 g    1307 t
ORIGIN
```

```
  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db     137 GACATTGTCGCAAGTCACAGAATTA 161
```

```
RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a    1552 c    1695 g    1307 t
ORIGIN
```

```
  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db     137 GACATTGTCGCAAGTCACAGAATTA 161
```

```
RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
```

```
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORF1 RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a    1552 c    1695 g    1307 t
ORIGIN
```

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db     137 GACATTGTCGCAAGTCACAGAATTA 161
```

```
RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a    1552 c    1695 g    1307 t
ORIGIN
```

Query Match 100.0%; Score 25; DB 6; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;

```
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db    137 GACATTGTCGCAAGTCACAGAATTA 161


RESULT  10
AR101809
LOCUS       AR101809     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db    137 GACATTGTCGCAAGTCACAGAATTA 161


RESULT  11
AR101810
LOCUS       AR101810     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db    137 GACATTGTCGCAAGTCACAGAATTA 161


RESULT  12
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu].  Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
```

```
CDS             complement(147. .1430)
                /gene="ybhC"
                /function="orf; Not classified"
                /note="f427; 98 pct identical to fragment YBHC_ECOLI
                SW:P46130 (300 aa) but contains 127 additional C-terminal
                residues"
                /codon_start=1
                /transl_table=11
                /product="putative pectinesterase"
                /protein_id="AAC73859.1"
                /db_xref="GI:1786989"
                /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
```

```
  Query Match          100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db    3513 GACATTGTCGCAAGTCACAGAATTA 3537
```

```
RESULT  13
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
```

```
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSE[illegible]ASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
```

```
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGETSRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match             100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db    9599 GACATTGTCGCAAGTCACAGAATTA 9623


RESULT  14
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
```

```
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase alpha subunit"
                     /protein_id="BAB34177.1"
                     /db_xref="GI:13360213"
                     /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                     GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                     LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                     AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                     AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                     LADIGEALKTVLK"
gene                 2781. .3686
                     /gene="ECs0755"
CDS                  2781. .3686
                     /gene="ECs0755"
                     /note="probable transcriptional regulator,similar to
                     transcriptional regulators e.g. glycine cleavage system
                     transcription activator (gcv operon activator) -
                     Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                     identity 31 in 300 aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative transcription regulator"
                     /protein_id="BAB34178.1"
                     /db_xref="GI:13360214"
                     /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                     EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                     PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                     LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                     NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                     EKSDSPHLQQFREWLLAKSVVPQEM"
gene                 complement(3720. .4322)
                     /gene="ECs0756"
CDS                  complement(3720. .4322)
                     /gene="ECs0756"
                     /note="probable cob(I)alamin adenosyltransferase,similar
                     to cob(I)alamin adenosyltransferases (corrinoid
                     adenosyltransferases) e.g. [Escherichia coli]
                     gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative cob(I)alamin adenosyltransferase"
                     /protein_id="BAB34179.1"
                     /db_xref="GI:13360215"
                     /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                     KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                     QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                     IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene                 complement(4332. .5984)
                     /gene="ECs0757"
CDS                  complement(4332. .5984)
                     /gene="ECs0757"
                     /note="probable fumarate hydratase,similar to fumarate
                     hydratases e.g. fumarate hydratase class I, aerobic
                     (fumarase) - Escherichia coli
                     gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative fumarate hydratase"
                     /protein_id="BAB34180.1"
                     /db_xref="GI:13360216"
                     /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                     ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                     NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                     TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                     GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                     EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                     EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                     DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                     QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                     MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene                 complement(6092. .7372)
                     /gene="ECs0758"
CDS                  complement(6092. .7372)
                     /gene="ECs0758"
                     /note="probable transport protein,similar to
                     glutamate/aspartate transport proteins (proton glutamate
                     symport proteins) e.g. [Bacillus stearothermophilus]
                     gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                     aa,also similar to C4-dicarboxylate transpor"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative transport protein"
                     /protein_id="BAB34181.1"
                     /db_xref="GI:13360217"
                     /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                     TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                     NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                     FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                     VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                     GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                     LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                     AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene                 complement(7533. .7850)
                     /gene="ECs0759"
CDS                  complement(7533. .7850)
                     /gene="ECs0759"
                     /note="unknown"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="BAB34182.1"
```

```
 Query Match           100.0%;  Score 25;  DB 1;  Length 297816;
 Best Local Similarity  100.0%;  Pred. No. 8.6e-06;
 Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db   92331 GACATTGTCGCAAGTCACAGAATTA 92355


RESULT  15
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
```

```
     source          1. .5526
                     /organism="uncultured bacterium pCosHE2"
                     /db_xref="taxon:143797"
                     /clone="pCosHE2"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     CDS             complement(52. .528)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical 17.1 kDa protein in modC-bioA
                     intergenic region"
                     /protein_id="AAG60577.1"
                     /db_xref="GI:12620125"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(587. .1876)
                     /gene="bioA"
     CDS             complement(587. .1876)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase BioA"
                     /protein_id="AAG60578.1"
                     /db_xref="GI:12620126"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            1963. .3003
                     /gene="bioB"
     CDS             1963. .3003
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG60579.1"
                     /db_xref="GI:12620127"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3000. .4154
                     /gene="bioF"
     CDS             3000. .4154
                     /gene="bioF"
                     /note="8-amino-7-oxononanoate synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA synthetase BioF"
                     /protein_id="AAG60580.1"
                     /db_xref="GI:12620128"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4141. .4896
                     /gene="bioC"
     CDS             4141. .4896
                     /gene="bioC"
                     /note="reaction step prior to synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60581.1"
                     /db_xref="GI:12620129"
                     /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                     HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4883. .>5526
                     /gene="bioD"
     CDS             4883. .>5526
                     /gene="bioD"
                     /note="DTB synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase BioD"
                     /protein_id="AAG60582.1"
                     /db_xref="GI:12620130"
                     /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                     GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                     SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                     TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT     1274 a    1507 c    1567 g    1178 t
ORIGIN


  Query Match             84.0%;  Score 21;  DB 1;  Length 5526;
  Best Local Similarity   100.0%;  Pred. No. 0.0052;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcacagaatta 25
           |||||||||||||||||||||
Db    1987 TTGTCGCAAGTCACAGAATTA 2007


RESULT  16
AF248314
LOCUS       AF248314     8227 bp    DNA              BCT       24-JAN-2001
DEFINITION  Uncultured bacterium pCosAS1 urocanase-like protein (hutU) gene,
            partial cds; histidine ammonia-lyase-like protein (hutH), DAPA
            aminotransferase BioA (bioA), biotin synthase BioB (bioB), 7-KAPA
            synthetase (bioF), biotin biosynthesis BioC-like protein (bioC),
            and dethiobiotin synthase BioD (bioD) genes, complete cds; ABC
            transporter-like protein (elsA) gene, partial cds; and unknown
            gene.
ACCESSION   AF248314
VERSION     AF248314.1  GI:12407610
KEYWORDS    .
SOURCE      uncultured bacterium pCosAS1.
  ORGANISM  uncultured bacterium pCosAS1
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 8227)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct Cloning from Enrichment Cultures, a Reliable Strategy for
            Isolation of Complete Operons and Genes from Microbial Consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
   PUBMED   11133432
REFERENCE   2  (bases 1 to 8227)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .8227
                     /organism="uncultured bacterium pCosAS1"
                     /db_xref="taxon:143793"
                     /clone="pCosAS1"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     gene            <1. .502
```

```
                /gene="hutU"
     CDS        <1. .502
                /gene="hutU"
                /note="putative urocanate hydratase"
                /codon_start=2
                /transl_table=11
                /product="urocanase-like protein"
                /protein_id="AAG53593.1"
                /db_xref="GI:12407618"
                /translation="PRERISFQGLPARICWVGLEWRQKLGLAFNEMVRRGEVSAPIVI
                GRDHLDSGSVASPNRETEAMQDGSDAVSDWPLLNALLNTASGATWVSLHHGGGVGMGF
                SQHAGMVIVCDGTDDAAARIARVLHNDPATGVMRHADAGYDIAIDCAKEQGLNLPMIA
                ATQGQH"
     gene       502. .2031
                /gene="hutH"
     CDS        502. .2031
                /gene="hutH"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="histidine ammonia-lyase-like protein"
                /protein_id="AAG53586.1"
                /db_xref="GI:12407611"
                /translation="MNALTLTPGTLTLAQLRQVWQQPLQLTLDESAHEAINDSVACVE
                AIVAEGRTAYGINTGFGLLAQTRIATHDLENLQRSLVLSHAAGVGEPLDDDIVRLMMV
                LKINSLARGFSGIRLSVIQALIALVNAGVYSVDPAKGSVGASGDLAPLAHMSLTLLGE
                GKARYRGEWLPAATALQKAGLAPVTLAAKEGLALLNGTQASTAFALRGLFEAEDLFAS
                AVVCGALTTEAVLGSRRPFDARIHEVRGQRGQIDAAALFRHVLTDTSAIAESHHNCDK
                VQDPYSLRCQPQVMGACLTQMRQVAEVLLVESNAVSDNPLVFAAENEMVFRGNFHAEP
                VAMAADNLALAIAEIGALSERRIALMMDKHMSQLPPFLVRNGGVNSGFMIAQVTAAAL
                ASENKGLCHPHTTSVDKPCRPSANQEDHVSMAPAAGRRLWEMAGNTRGVLAVEWLAAC
                QGADLRDGLTSSPLLEQARQSCGEQVAHYDDDRFFAPDIEAAISLLNKGSLVGLLPAF
                L"
     CDS        complement(2130. .2606)
                /note="ORF1; similar to hypothetical 17.1 kDa protein in
                modC-bioA intergenic region"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAG53587.1"
                /db_xref="GI:12407612"
                /translation="MKLISQDLRDGDKLPLRHVFNGMGYEGDNISPHLAWDEVPSGTK
                SFVVTCYDPDAPTDSGWWHWIVANIPADTRVLPQGSGSSLVALPEGAVETRTDFGKAG
                YGGAAPPKGETHRYIFTVHALNVERIEVDEEASGAMVGFNVHFHTLGSASITAMYS"
     gene       complement(2651. .3940)
                /gene="bioA"
     CDS        complement(2651. .3940)
                /gene="bioA"
                /note="7,8-diamino-pelargonic acid aminotransferase"
                /codon_start=1
                /transl_table=11
                /product="DAPA aminotransferase BioA"
                /protein_id="AAG53588.1"
                /db_xref="GI:12407613"
                /translation="MTTDDLAFDRQHIWHPYTSMTQPLPVYPVVEAHGCELTLASGEQ
                LVDGMSSWWAAIHGYNHPRLNAAMKAQIDRMSHVMFGGITHQPAVDLCRRLVAMTPET
                LECVFPADSGSVAVEVAMKMALQYWHAKGESRQRFITFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRFDGAWDEMDMVGLAPLMAAHRHEIAAVILEPIVQGAG
                GMRMYHPEWLKRIRRMCDREGILLIADEIATGFGRTGKLFACEHAGITPDILCLGKAL
                TGGTMTMSATLTTRQVAETISNGEAGCFMHGPTFMGNPLACAVASESLALLESGEWQD
                QVAAIESQLQEELAVARDSEFVADVRVLGAIGVVETTHPVNMAALQRFFVARGVWVRP
                FGKLIYLMPPYVIRPDQLSKLTRAVVDAVQAPAHFTR"
     gene       4029. .5069
                /gene="bioB"
     CDS        4029. .5069
                /gene="bioB"
                /codon_start=1
                /transl_table=11
                /product="biotin synthase BioB"
                /protein_id="AAG53589.1"
                /db_xref="GI:12407614"
                /translation="MAHQARWTLSQVTALFEKPLLELLFEAQQIHRQHFDPQQIQVST
                LLSIKTGACPEDCKYCPQSARYKTGLESERLMEVEQVLESARQAKNAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLQACMTLGTLDDHQAQRLASAGLDYYNHNLDTFPEF
                YGNIITTRTYQERLDTLDKVRDAGIKVCSGGIVGLCETVTDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVXAFDXIRTXSVAXVMXPTXLVRLSTGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDVQLFRKLGLNPQQTAVLTGDNEQQHQLEQQLFN
                ADTDQFYNAATV"
     gene       5066. .6223
                /gene="bioF"
     CDS        5066. .6223
                /gene="bioF"
                /note="7-keto-8-amino-pelargonic acid synthetase; similar
                to BioF"
                /codon_start=1
                /transl_table=11
                /product="7-KAPA synthetase"
                /protein_id="AAG53590.1"
                /db_xref="GI:12407615"
                /translation="MTWQQRIDAALDARRAADALRVRTPVENGAGRFLIREQRRYCNF
                SGNDYLGLSQHPEIIRAWQQGAEQYGVGSGGSGHVSGYTHAHQALEEHLAGWLGYPRA
                LLFISGFAANQALIAALTDKEDRIVADRLSHASLLEAASHSPAQLRRFAHNDVAQLDS
                LLDKPCHGQQLVVTEGIFSMDGDSAPLGVIHAAAQRQGSWLLVDDAHGIGVTGPEGRG
                SAASQGVKPELLVVTFGKGVGVSGAAVLCSDAVADYLLQFARHLIYSTSMPPAQAVAL
                LAAFNVIRSDEDNARRQRLAALIDQFRAGVRDLPVQITASQSAIQPLIVGENARALHL
                AQRLRERGIWVTAIRPPTVPAGTARLRLTLTAAHQPADITALVEALHVAGE"
     gene       6207. .6962
                /gene="bioC"
     CDS        6207. .6962
                /gene="bioC"
                /codon_start=1
                /transl_table=11
                /product="biotin biosynthesis BioC-like protein"
                /protein_id="AAG53591.1"
                /db_xref="GI:12407616"
                /translation="MSQVNKQAIAAAFGRAAQSYSRHDALQRQSADGLRALLAENTFA
                SVLDAGCGPGSNSLAWRQTGSVVTAMDLCAPMLEEARRNQAADSYLQADIEAIPLADA
                QFSLAWSHLAVQXCASLPRALAELWRVTQLGGKVAFTTLLNGSLPELNQAWQAVDTQP
                HANRFLSGQQVNDALSGWQYRCAVQTVTLLFDDALSAMHSLKGIGATHLHAGRAAKPL
                TRGQLQQLAQAWPQQAGKFPLSYQLFHGIIERD"
     gene       6940. .7680
                /gene="bioD"
     CDS        6940. .7680
                /gene="bioD"
                /codon_start=1
                /transl_table=11
                /product="dethiobiotin synthase BioD"
                /protein_id="AAG53592.1"
                /db_xref="GI:12407617"
                /translation="MELSNVTERYFVTGTDTEVGKTVASTALLQAARALGRHTAGYKP
                VASGSEMTAQGLRNSDALALQRNSSLALSYPAVNPYTFAEPTSPHIVSADEQRPIEFA
                TLSAGLRALETQADWVLVEGAGGWFTPLSQEQTFADWVMAERLPVILVVGVKLGCINH
                AMLTAQAIRQAGLHFAGWVANGVVVPGKRHAEYMATLRRVLSAPLLGEIPWLGEQADT
                SAVGQYLDLTALTPSAPSSGQSAGHPAR"
     gene       complement(7619. .>8227)
                /gene="elsA"
     CDS        complement(7619. .>8227)
                /gene="elsA"
                /note="putative urea/short-chain amide ABC transporter"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter-like protein"
                /protein_id="AAG53594.1"
                /db_xref="GI:12407619"
                /translation="DCXLRTRXGENXLVSCIAGLLPIASGTMVWHEAPGAVNQAGTEI
                XVVPQDRRIFSQLTVEENLHIAQMAKGIAGGSLSRDVFEFFPELYPLRQLRASGLNDD
                QQYQLALANALVTRPRLLILDEPSRGTGQLFTVKLAQLLLRFNRELGLSVLLAEQHLS
                LIRRVADRFCLLYRGRNVAQGRVSELDDPLIAHWMGPTASAQ"
BASE COUNT     1512 a    2642 c    2605 g    1453 t      15 others
ORIGIN


  Query Match             72.0%;  Score 18;  DB 1;  Length 8227;
  Best Local Similarity   100.0%;  Pred. No. 0.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 gacattgtcgcaagtcac 18
           ||||||||||||||||||
Db    4049 GACATTGTCGCAAGTCAC 4066


RESULT  17
AC087305/c
LOCUS       AC087305    32767 bp    DNA             HTG       23-DEC-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-238P13 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC087305
VERSION     AC087305.1  GI:11990696
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 32767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-238P13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 32767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12071
                Center clone name: 238_P_13
            --------------.
            * NOTE: This record contains 40 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      707: contig of 707 bp in length
            *      708 807: gap of      100 bp
            *      808     1527: contig of 720 bp in length
            *     1528 1627: gap of      100 bp
            *     1628     2348: contig of 721 bp in length
            *     2349 2448: gap of      100 bp
            *     2449     3157: contig of 709 bp in length
            *     3158 3257: gap of      100 bp
            *     3258     3973: contig of 716 bp in length
            *     3974 4073: gap of      100 bp
            *     4074     4799: contig of 726 bp in length
            *     4800 4899: gap of      100 bp
            *     4900     5617: contig of 718 bp in length
            *     5618 5717: gap of      100 bp
            *     5718     6450: contig of 733 bp in length
            *     6451 6550: gap of      100 bp
            *     6551     7259: contig of 709 bp in length
            *     7260 7359: gap of      100 bp
            *     7360     8066: contig of 707 bp in length
            *     8067 8166: gap of      100 bp
            *     8167     8907: contig of 741 bp in length
            *     8908 9007: gap of      100 bp
            *     9008     9716: contig of 709 bp in length
            *     9717 9816: gap of      100 bp
            *     9817    10532: contig of 716 bp in length
            *    10533 10632: gap of      100 bp
            *    10633    11354: contig of 722 bp in length
            *    11355 11454: gap of      100 bp
            *    11455    12166: contig of 712 bp in length
            *    12167 12266: gap of      100 bp
            *    12267    12976: contig of 710 bp in length
            *    12977 13076: gap of      100 bp
            *    13077    13789: contig of 713 bp in length
            *    13790 13889: gap of      100 bp
            *    13890    14617: contig of 728 bp in length
            *    14618 14717: gap of      100 bp
            *    14718    15452: contig of 735 bp in length
            *    15453 15552: gap of      100 bp
            *    15553    16276: contig of 724 bp in length
            *    16277 16376: gap of      100 bp
            *    16377    17111: contig of 735 bp in length
            *    17112 17211: gap of      100 bp
            *    17212    17925: contig of 714 bp in length
            *    17926 18025: gap of      100 bp
            *    18026    18746: contig of 721 bp in length
            *    18747 18846: gap of      100 bp
            *    18847    19569: contig of 723 bp in length
            *    19570 19669: gap of      100 bp
            *    19670    20384: contig of 715 bp in length
            *    20385 20484: gap of      100 bp
            *    20485    21190: contig of 706 bp in length
            *    21191 21290: gap of      100 bp
            *    21291    22028: contig of 738 bp in length
            *    22029 22128: gap of      100 bp
            *    22129    22841: contig of 713 bp in length
            *    22842 22941: gap of      100 bp
            *    22942    23653: contig of 712 bp in length
            *    23654 23753: gap of      100 bp
            *    23754    24478: contig of 725 bp in length
            *    24479 24578: gap of      100 bp
            *    24579    25305: contig of 727 bp in length
            *    25306 25405: gap of      100 bp
            *    25406    26143: contig of 738 bp in length
            *    26144 26243: gap of      100 bp
            *    26244    26968: contig of 725 bp in length
            *    26969 27068: gap of      100 bp
            *    27069    27785: contig of 717 bp in length
            *    27786 27885: gap of      100 bp
            *    27886    28584: contig of 699 bp in length
            *    28585 28684: gap of      100 bp
            *    28685    29467: contig of 783 bp in length
            *    29468 29567: gap of      100 bp
            *    29568    30292: contig of 725 bp in length
            *    30293 30392: gap of      100 bp
            *    30393    31117: contig of 725 bp in length
            *    31118 31217: gap of      100 bp
            *    31218    31944: contig of 727 bp in length
```

```
            *    31945 32044: gap of      100 bp
            *    32045    32767: contig of 723 bp in length.
FEATURES             Location/Qualifiers
     source          1. .32767
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-238P13"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT     8045 a   5670 c   5396 g   9330 t   4326 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 32767;
  Best Local Similarity   100.0%;  Pred. No. 6.3;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatta 25
           ||||||||||||||||
Db  30730 GCAAGTCACAGAATTA 30715


RESULT  18
AC021110
LOCUS       AC021110   155172 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-238P13, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC021110
VERSION     AC021110.3  GI:8439978
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155172)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155172)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 10, 2000 this sequence version replaced gi:7021826.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0238P13
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 73%
            Sequencing vector: plasmid; 27%
            Chemistry: Dye-primer ET; 73% of reads
            Chemistry: Dye-terminator Big Dye; 27% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 151162 bases at least Q40
            Consensus quality: 152443 bases at least Q30
            Consensus quality: 153249 bases at least Q20
            Insert size: 177000; agarose-fp
            Insert size: 154672; sum-of-contigs
            Quality coverage: 4.08 in Q20 bases; agarose-fp
            Quality coverage: 4.69 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9265: contig of 9265 bp in length
            *     9266     9365: gap of unknown length
            *     9366    26504: contig of 17139 bp in length
            *    26505    26604: gap of unknown length
            *    26605    44300: contig of 17696 bp in length
            *    44301    44400: gap of unknown length
            *    44401    62250: contig of 17850 bp in length
            *    62251    62350: gap of unknown length
            *    62351   105744: contig of 43394 bp in length
            *   105745   105844: gap of unknown length
            *   105845   155172: contig of 49328 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155172
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /clone="RP11-238P13"
     misc_feature    1. .9265
                     /note="assembly_name:Contig1"
     misc_feature    9366. .26504
                     /note="assembly_name:Contig2"
     misc_feature    26605. .44300
                     /note="assembly_name:Contig3
                     clone_end:T7
                     vector_side:right"
     misc_feature    44401. .62250
                     /note="assembly_name:Contig4"
     misc_feature    62351. .105744
                     /note="assembly_name:Contig5"
     misc_feature    105845. .155172
                     /note="assembly_name:Contig6"
BASE COUNT     51543 a  28385 c  28043 g  46698 t    503 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 155172;
  Best Local Similarity   100.0%;  Pred. No. 5.1;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatta 25
           ||||||||||||||||
Db 111929 GCAAGTCACAGAATTA 111944


RESULT  19
AC041018/c
LOCUS       AC041018   156184 bp    DNA             HTG       22-MAY-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-478O20 map 4, WORKING DRAFT
            SEQUENCE, 16 unordered pieces.
ACCESSION   AC041018
VERSION     AC041018.2  GI:8016453
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156184)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-478O20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156184)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
```

```
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 22, 2000 this sequence version replaced gi:7534220.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9577
                Center clone name: 478_O_20
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 147958 bases at least Q40
                Consensus quality: 151874 bases at least Q30
                Consensus quality: 153413 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 154684; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1351: contig of 1351 bp in length
            *     1352 1451: gap of      100 bp
            *     1452     4369: contig of 2918 bp in length
            *     4370 4469: gap of      100 bp
            *     4470     8873: contig of 4404 bp in length
            *     8874 8973: gap of      100 bp
            *     8974    13167: contig of 4194 bp in length
            *    13168 13267: gap of      100 bp
            *    13268    18989: contig of 5722 bp in length
            *    18990 19089: gap of      100 bp
            *    19090    24989: contig of 5900 bp in length
            *    24990 25089: gap of      100 bp
            *    25090    33460: contig of 8371 bp in length
            *    33461 33560: gap of      100 bp
            *    33561    42328: contig of 8768 bp in length
            *    42329 42428: gap of      100 bp
            *    42429    50674: contig of 8246 bp in length
            *    50675 50774: gap of      100 bp
            *    50775    59862: contig of 9088 bp in length
            *    59863 59962: gap of      100 bp
            *    59963    71699: contig of 11737 bp in length
            *    71700 71799: gap of      100 bp
            *    71800    82583: contig of 10784 bp in length
            *    82584 82683: gap of      100 bp
            *    82684    95072: contig of 12389 bp in length
            *    95073 95172: gap of      100 bp
            *    95173   108949: contig of 13777 bp in length
            *   108950 109049: gap of      100 bp
            *   109050   126668: contig of 17619 bp in length
            *   126669 126768: gap of      100 bp
            *   126769   156184: contig of 29416 bp in length.
  FEATURES             Location/Qualifiers
       source          1. .156184
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /chromosome="4"
                       /map="4"
                       /clone="RP11-478O20"
                       /clone_lib="RPCI-11 Human Male BAC"
       misc_feature    1. .1351
                       /note="assembly_fragment"
       misc_feature    1452. .4369
                       /note="assembly_fragment"
       misc_feature    4470. .8873
                       /note="assembly_fragment
                       clone_end:SP6
                       vector_side:right"
       misc_feature    8974. .13167
                       /note="assembly_fragment"
       misc_feature    13268. .18989
                       /note="assembly_fragment"
       misc_feature    19090. .24989
                       /note="assembly_fragment"
       misc_feature    25090. .33460
                       /note="assembly_fragment"
       misc_feature    33561. .42328
                       /note="assembly_fragment"
       misc_feature    42429. .50674
                       /note="assembly_fragment"
       misc_feature    50775. .59862
                       /note="assembly_fragment"
       misc_feature    59963. .71699
                       /note="assembly_fragment"
       misc_feature    71800. .82583
                       /note="assembly_fragment"
       misc_feature    82684. .95072
                       /note="assembly_fragment
                       clone_end:T7
                       vector_side:right"
       misc_feature    95173. .108949
                       /note="assembly_fragment"
       misc_feature    109050. .126668
                       /note="assembly_fragment"
       misc_feature    126769. .156184
                       /note="assembly_fragment"
  BASE COUNT    51507 a  26941 c  26653 g  49578 t   1505 others
  ORIGIN


    Query Match             64.0%;  Score 16;  DB 2;  Length 156184;
    Best Local Similarity   100.0%;  Pred. No. 5.1;
    Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  Qy      10 gcaagtcacagaatta 25
             ||||||||||||||||
  Db 123532 GCAAGTCACAGAATTA 123517


  RESULT  20
  AP001390/c
  LOCUS       AP001390 ' 174815 bp    DNA             HTG       30-MAY-2000
  DEFINITION  Homo sapiens chromosome 18 clone RP11-635B11 map 18q23, WORKING
              DRAFT SEQUENCE, 21 unordered pieces.
  ACCESSION   AP001390
  VERSION     AP001390.2  GI:8117301
  KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
  SOURCE      Homo sapiens DNA, clone:RP11-635B11.
    ORGANISM  Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 174815)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 174,815 genomic DNA of 18q23
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 174815)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7259325.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-635B11
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 159194 bases at least Q40
                Consensus quality: 167087 bases at least Q30
                Consensus quality: 170693 bases at least Q20
                Insert size: 172815; sum-of-contigs
                Quality coverage: 4.94x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            21 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     58823 contig of  58823 bp in length
              58924     73243 contig of  14320 bp in length
              73344     83572 contig of  10229 bp in length
              83673     94547 contig of  10875 bp in length
              94648    104597 contig of   9950 bp in length
             104698    114347 contig of   9650 bp in length
             114448    122080 contig of   7633 bp in length
             122181    128465 contig of   6285 bp in length
             128566    135791 contig of   7226 bp in length
             135892    142934 contig of   7043 bp in length
             143035    147999 contig of   4965 bp in length
             148100    151734 contig of   3635 bp in length
             151835    155576 contig of   3742 bp in length
             155677    159193 contig of   3517 bp in length
             159294    161359 contig of   2066 bp in length
             161460    163515 contig of   2056 bp in length
             163616    166614 contig of   2999 bp in length
             166715    169056 contig of   2342 bp in length
             169157    171527 contig of   2371 bp in length
             171628    173248 contig of   1621 bp in length
             173349    174815 contig of   1467 bp in length
            Sequence updated (26-May-2000).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    58823: contig of 58823 bp in length
            *    58824 58923: gap of      100 bp
            *    58924    73243: contig of 14320 bp in length
            *    73244 73343: gap of      100 bp
            *    73344    83572: contig of 10229 bp in length
            *    83573 83672: gap of      100 bp
            *    83673    94547: contig of 10875 bp in length
            *    94548 94647: gap of      100 bp
            *    94648   104597: contig of 9950 bp in length
            *   104598 104697: gap of      100 bp
            *   104698   114347: contig of 9650 bp in length
            *   114348 114447: gap of      100 bp
            *   114448   122080: contig of 7633 bp in length
            *   122081 122180: gap of      100 bp
            *   122181   128465: contig of 6285 bp in length
            *   128466 128565: gap of      100 bp
            *   128566   135791: contig of 7226 bp in length
            *   135792 135891: gap of      100 bp
            *   135892   142934: contig of 7043 bp in length
            *   142935 143034: gap of      100 bp
            *   143035   147999: contig of 4965 bp in length
            *   148000 148099: gap of      100 bp
            *   148100   151734: contig of 3635 bp in length
            *   151735 151834: gap of      100 bp
            *   151835   155576: contig of 3742 bp in length
            *   155577 155676: gap of      100 bp
            *   155677   159193: contig of 3517 bp in length
            *   159194 159293: gap of      100 bp
            *   159294   161359: contig of 2066 bp in length
            *   161360 161459: gap of      100 bp
            *   161460   163515: contig of 2056 bp in length
            *   163516 163615: gap of      100 bp
            *   163616   166614: contig of 2999 bp in length
            *   166615 166714: gap of      100 bp
            *   166715   169056: contig of 2342 bp in length
            *   169057 169156: gap of      100 bp
            *   169157   171527: contig of 2371 bp in length
            *   171528 171627: gap of      100 bp
            *   171628   173248: contig of 1621 bp in length
            *   173249 173348: gap of      100 bp
            *   173349   174815: contig of 1467 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174815
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q23"
                     /clone="RP11-635B11"
     misc_feature    1. .58823
                     /note="assembly_fragment"
     misc_feature    58924. .73243
                     /note="assembly_fragment"
     misc_feature    73344. .83572
                     /note="assembly_fragment"
     misc_feature    83673. .94547
                     /note="assembly_fragment"
     misc_feature    94648. .104597
                     /note="assembly_fragment"
     misc_feature    104698. .114347
                     /note="assembly_fragment"
     misc_feature    114448. .122080
                     /note="assembly_fragment"
     misc_feature    122181. .128465
                     /note="assembly_fragment"
     misc_feature    128566. .135791
                     /note="assembly_fragment"
     misc_feature    135892. .142934
                     /note="assembly_fragment"
     misc_feature    143035. .147999
                     /note="assembly_fragment"
     misc_feature    148100. .151734
                     /note="assembly_fragment"
     misc_feature    151835. .155576
                     /note="assembly_fragment"
     misc_feature    155677. .159193
```

```
                     /note="assembly_fragment"
     misc_feature    159294. .161359
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    161460. .163515
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    163616. .166614
                     /note="assembly_fragment"
     misc_feature    166715. .169056
                     /note="assembly_fragment"
     misc_feature    169157. .171527
                     /note="assembly_fragment"
     misc_feature    171628. .173248
                     /note="assembly_fragment"
     misc_feature    173349. .174815
                     /note="assembly_fragment"
BASE COUNT    52310 a  30923 c  32183 g  57399 t   2000 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 174815;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatta 25
           ||||||||||||||||
Db   44384 GCAAGTCACAGAATTA 44369


RESULT  21
AC091589/c
LOCUS       AC091589   178056 bp    DNA              HTG       26-JUL-2001
DEFINITION  Homo sapiens chromosome 18 clone RP11-635B11 map 18, *** SEQUENCING
            IN PROGRESS ***, 2 ordered pieces.
ACCESSION   AC091589
VERSION     AC091589.6  GI:15022066
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178056)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-635B11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178056)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 26, 2001 this sequence version replaced gi:14994291.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ------------- Project Information
                Center project name: L13182
                Center clone name: 635_B_11
            -------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   168134: contig of 168134 bp in length
            *  168135 168234: gap of      100 bp
            *  168235   178056: contig of 9822 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178056
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-635B11"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    49585 a  30884 c  34094 g  63393 t    100 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 178056;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatta 25
           ||||||||||||||||
Db   64175 GCAAGTCACAGAATTA 64160


RESULT  22
AC020734/c
LOCUS       AC020734   196550 bp    DNA              PRI       09-AUG-2001
DEFINITION  Homo sapiens chromosome 4 clone RP11-665I14, complete sequence.
ACCESSION   AC020734
VERSION     AC020734.6  GI:15145620
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196550)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196550)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 196550)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 9, 2001 this sequence version replaced gi:15022773.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
```

```
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------- Project Information ----------
            Center project name: H_NH0665I14
            ----------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .196550
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-665I14"
BASE COUNT    65000 a  34953 c  34110 g  62487 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 9;  Length 196550;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatta 25
           ||||||||||||||||
Db   55717 GCAAGTCACAGAATTA 55702


RESULT  23
AX136747
LOCUS       AX136747      561 bp    DNA             PAT       30-MAY-2001
DEFINITION  Sequence 669 from Patent EP1067182.
ACCESSION   AX136747
VERSION     AX136747.1  GI:14273151
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
            Hayashi,K.
  TITLE     Secretory protein or membrane protein
  JOURNAL   Patent: EP 1067182-A 669 10-JAN-2001;
            Helix Research Institute (JP)
FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      213 a     83 c    114 g    149 t      2 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 6;  Length 561;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db    339 CAAGTCACAGAATTA 353


RESULT  24
AR080743
LOCUS       AR080743     1771 bp    DNA             PAT       31-AUG-2000
DEFINITION  Sequence 48 from patent US 5968826.
ACCESSION   AR080743
VERSION     AR080743.1  GI:10007473
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1771)
  AUTHORS   Bennett,C.Frank, Condon,T.P. and Cowsert,L.M.
  TITLE     Antisense inhibition of integrin .alpha.4 expression
  JOURNAL   Patent: US 5968826-A 48 19-OCT-1999;
FEATURES             Location/Qualifiers
     source          1. .1771
                     /organism="unknown"
BASE COUNT      371 a    471 c    519 g    410 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 6;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db    312 GCAAGTCACAGAATT 326


RESULT  25
MUSINTESUB
LOCUS       MUSINTESUB   1771 bp    DNA             ROD       18-APR-1996
DEFINITION  Mus musculus (clone A4G16) integrin alpha-4 subunit gene, 5' end
            cds.
ACCESSION   L20788
VERSION     L20788.1  GI:309416
KEYWORDS    integrin alpha-4 subunit.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1771)
  AUTHORS   De Meirsman,C., Schollen,E., Jaspers,M., Ongena,K., Matthijs,G.,
            Marynen,P. and Cassiman,J.J.
  TITLE     Cloning and characterization of the promoter region of the murine
            alpha-4 integrin subunit
  JOURNAL   DNA Cell Biol. 13 (7), 743-754 (1994)
  MEDLINE   95290094
FEATURES             Location/Qualifiers
     source          1. .1771
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="3T3"
                     /cell_type="fibroblasts"
                     /tissue_lib="NIH 3T3 in lambda-FIXII"
     GC_signal       503. .508
                     /note="putative"
     GC_signal       845. .850
     exon            952. .1389
     GC_signal       complement(1055. .1060)
     gene            join(1193. .1389,1711. .1771)
                     /gene="VLA-4"
     CDS             join(1193. .1389,1711. .>1771)
                     /gene="VLA-4"
                     /note="putative"
                     /codon_start=1
                     /product="integrin alpha-4 subunit"
                     /protein_id="AAA97501.1"
                     /db_xref="GI:309417"
                     /translation="MAAEARCRPRSRGIALREAVMLLLYFGVPTGHSYNLDPENALLY
                     QGPSGTLFGYSVVLHSHGSKRWLIVGAPTASWLSNASVVNPG"
     intron          1390. .1710
     exon            1711. .1771
                     /gene="VLA-4"
BASE COUNT      371 a    471 c    519 g    410 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db    312 GCAAGTCACAGAATT 326
```

```
RESULT  26
G07181/c
LOCUS       G07181       1829 bp    DNA             STS       19-OCT-1995
DEFINITION  human STS WI-9211.
ACCESSION   G07181
VERSION     G07181.1  GI:860426
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
            Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1829)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped ESTs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: CAGCCACACTAACTGCATTAGC
            Primer B: GACAGAAATGGATTGTATAAATGGC
            STS size: 111
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
                Taq Polymerase: 0.025 units/ul
                Total Vol: 20 ul

            Buffer:
                MgCl2: 1.5 mM
                KCl: 50 mM
                Tris-HCL: 10 mM
                pH: 9.3

            Prepared with primer pairs derived from D13644 -- Unigene.
FEATURES                Location/Qualifiers
     source             1. .1829
                        /organism="Homo sapiens"
     STS                891. .1001
     primer_bind        891. .912
     primer_bind        complement(977. .1001)
BASE COUNT      539 a    321 c    320 g    590 t     59 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 1829;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db    1689 CAAGTCACAGAATTA 1675


RESULT  27
```

```
MMINTEGR01
LOCUS       MMINTEGR01   1949 bp    DNA             ROD       02-OCT-1996
DEFINITION  Mus musculus alpha-4 integrin gene, exons 1 and 2.
ACCESSION   U34627
VERSION     U34627.1  GI:1173582
KEYWORDS    .
SEGMENT     1 of 22
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   De Meirsman,C., Schollen,E., Jaspers,M., Ongena,K., Matthijs,G.,
            Marynen,P. and Cassiman,J.J.
  TITLE     Cloning and characterization of the promoter region of the murine
            alpha-4 integrin subunit
  JOURNAL   DNA Cell Biol. 13 (7), 743-754 (1994)
  MEDLINE   95290094
REFERENCE   2  (bases 1 to 1949)
  AUTHORS   De Meirsman,C., Jaspers,M., Schollen,E. and Cassiman,J.J.
  TITLE     The genomic structure of the murine alpha 4 integrin gene
  JOURNAL   DNA Cell Biol. 15 (7), 595-603 (1996)
  MEDLINE   96326295
REFERENCE   3  (bases 1 to 1949)
  AUTHORS   De Meirsman,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-1995) Catherine De Meirsman, Center for Human
            Genetics, University of Leuven, Herestraat 49, Leuven 3000, Belgium
FEATURES                Location/Qualifiers
     source             1. .1949
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /tissue_type="spleen"
                        /note="NIH-3T3 B6/CBA spleen genomic library in lambda Fix
                        II; NIH-3T3 nd B6/CBA spleen genomic library in lambda
                        FixII"
     promoter           1. .952
     exon               953. .1389
                        /number=1
     5'UTR              953. .1192
     exon               1711. .1832
                        /number=2
BASE COUNT      420 a    514 c    563 g    452 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 1949;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db    312 GCAAGTCACAGAATT 326


RESULT  28
ECHK31IM
LOCUS       ECHK31IM     2340 bp    DNA             BCT       25-SEP-1998
DEFINITION  E.coli HK31IM gene.
ACCESSION   X82231
VERSION     X82231.1  GI:3660492
KEYWORDS    alpha-polypeptide; beta-polypeptide; R.EcoHK31I protein.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 2340)
  AUTHORS   Lee,K.F., Kam,K.M. and Shaw,P.C.
  TITLE     A bacterial methyltransferase M.EcoHK311 requires two proteins for
            in vitro methylation
  JOURNAL   Nucleic Acids Res. 23 (1), 103-108 (1995)
  MEDLINE   95175351
REFERENCE   2  (bases 1 to 2340)
```

```
  AUTHORS   Shaw,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-1994) P. Shaw, Dept. of Biochemistry, The Chinese
            University of Hong Kong, Shatin NT, HONG KONG
  REMARK    revised by [3]
REFERENCE   3  (bases 1 to 2340)
  AUTHORS   Shaw,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1997) P. Shaw, Dept. of Biochemistry, The Chinese
            University of Hong Kong, Shatin NT, HONG KONG
REFERENCE   4  (bases 1 to 2340)
  AUTHORS   Lee,K.F., Shaw,P.C., Picone,S.J., Wilson,G.G. and Lunnen,K.D.
  TITLE     Sequence comparison of the EcoHK31I and EaeI
            restriction-modification systems suggests an intergenic transfer of
            genetic material
  JOURNAL   Biol. Chem. 379 (4-5), 437-441 (1998)
  MEDLINE   98290299
COMMENT     On Sep 28, 1998 this sequence version replaced gi:639971.
FEATURES             Location/Qualifiers
     source          1. .2340
                     /organism="Escherichia coli"
                     /strain="HK31"
                     /db_xref="taxon:562"
     gene            407. .1336
                     /gene="ecoHK31IM alpha"
     CDS             407. .1336
                     /gene="ecoHK31IM alpha"
                     /function="methyltransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="M.EcoHK31I alpha polypeptide"
                     /protein_id="CAA57706.1"
                     /db_xref="GI:3660493"
                     /db_xref="SPTREMBL:O87740"
                     /translation="MKKKPLKQYKVSSFFAG1GGFDLGLEKAGMEVVFQCEINKFCQK
                     VLRKNWSKVPLHTDITRLNADEIPESNVWCGGFPCQDVSSANQGKRKGLEGARSGLFY
                     TYAKLIEERKPEWLIIENVPGLLNSHNGQDFKVVIDTLVEFGYGVSWRVLDAKYFGTP
                     QRRRRVYIVASLGDMRSARVLFEPGATRIVDKQGQGERANASGASATGLSKADIYSIQ
                     HASIGRKATAGPQAKGYRNDGETYTLDSRGSSDAVCAPDAAFRVRASTGISRELDSNR
                     FRAVGNAVAVPIVEWLGKRIVLVDQQVLNEALS"
     gene            717. .1247
                     /gene="ecoHK31IM beta"
     CDS             717. .1247
                     /gene="ecoHK31IM beta"
                     /function="methyltransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="M.EcoHK31I polypeptide beta"
                     /protein_id="CAA57705.1"
                     /db_xref="GI:3660494"
                     /db_xref="SPTREMBL:Q47257"
                     /translation="MQNSSKKESLNGLLLKMFPDCSTATMDKTSKLSSIRWSNSGMAF
                     RGEYWMQNTLEHPSVEEECTLSQVLETCAPLESFLNPEQLESLINRAKERGQMLPEPL
                     LQAYQKQISILSSMQVLDEKQPQDLKQKDTGTMEKPTLLTQEEVQMLYVRRMLPSEYE
                     RLQGFPENWTLIDSEQ"
     gene            complement(1317. .2273)
                     /gene="ecoHK31IR"
     CDS             complement(1317. .2273)
                     /gene="ecoHK31IR"
                     /function="restriction endonuclease"
                     /codon_start=1
                     /transl_table=11
                     /product="R.EcoHK31I protein"
                     /protein_id="CAA57707.1"
                     /db_xref="GI:3660495"
                     /db_xref="SPTREMBL:O87741"
                     /translation="MAVSIAEYLGQRTDVDQQIVPVAKGTQIQCPFMDRTCDKASKVK
                     NPTPPVCSVRKPDGTVWIVCEHRLCSTRQKKTVIVNGRKKQIENILVEHQRDILRKVA
                     KLIYQDPELQDSEIGVRREVNIPLPDSDNSYHADYVMRNFSGRGRVDEVLLEMQGGGE
                     TSSTGEITRHIAAWADLEFPTNEILRQPVAANTIETNAWRRQQEQFLVKGNVVDQTGG
                     KIVFAVGSLLYDYLHRRFRNANLRDLKAHNWTLCILAFKEDTSDEPRPGPIPLIIDDS
                     KTLFTNYSTFVRFLTDQGAPRPELFEGSFLRLDNSSVTIMPR"
BASE COUNT      679 a    458 c    527 g    676 t
```

```
ORIGIN


  Query Match             60.0%;  Score 15;  DB 1;  Length 2340;
  Best Local Similarity   100.0%;  Pred. No. 39;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 acattgtcgcaagtc 16
           ||||||||||||||||
Db     906 ACATTGTCGCAAGTC 920


RESULT  29
Y029487S18
LOCUS       Y029487S18   3140 bp    DNA             PRI       06-JUL-2001
DEFINITION  Homo sapiens aminophospholipid-transporting ATPase (ATP10C) gene,
            exons 20, 21, and complete cds.
ACCESSION   AY029504
VERSION     AY029504.1  GI:13878298
KEYWORDS    .
SEGMENT     18 of 18
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3140)
  AUTHORS   Herzing,L.B.K., Kim,S.J., Cook,E.H. Jr. and Ledbetter,D.H.
  TITLE     The human aminophospholipid-transporting ATPase gene ATP10C maps
            adjacent to UBE3A and exhibits similar imprinted expression
  JOURNAL   Am. J. Hum. Genet. 68 (6), 1501-1505 (2001)
  MEDLINE   21313119
   PUBMED   11353404
REFERENCE   2  (bases 1 to 3140)
  AUTHORS   Herzing,L.B.K., Kim,S.-J., Cook,E.H. Jr. and Ledbetter,D.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2001) Human Genetics and Psychiatry, University
            of Chicago, 920 E. 58th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .3140
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     mRNA            join(AY029487.1:1. .582,AY029488.1:96. .300,
                     AY029489.1:118. .203,AY029490.1:1. .107,AY029491.1:34. .165,
                     AY029492.1:713. .843,AY029493.1:58. .310,AY029494.1:88. .305,
                     AY029495.1:75. .269,AY029496.1:102. .669,
                     AY029497.1:185. .288,AY029497.1:383. .509,
                     AY029498.1:99. .283,AY029499.1:88. .415,AY029500.1:145. .221,
                     AY029501.1:119. .244,AY029502.1:148. .348,
                     AY029503.1:177. .257,AY029503.1:337. .441,105. .292,
                     439. .3140)
                     /gene="ATP10C"
                     /product="aminophospholipid-transporting ATPase"
     gene            order(AY029487.1:1. .1040,AY029488.1:1. .331,
                     AY029489.1:1. .357,AY029490.1:1. .107,AY029491.1:1. .333,
                     AY029492.1:1. .980,AY029493.1:1. .436,AY029494.1:1. .438,
                     AY029495.1:1. .396,AY029496.1:1. .743,AY029497.1:1. .588,
                     AY029498.1:1. .421,AY029499.1:1. .558,AY029500.1:1. .327,
                     AY029501.1:1. .269,AY029502.1:1. .459,AY029503.1:1. .487,
                     1. .3140)
                     /gene="ATP10C"
     CDS             join(AY029487.1:134. .582,AY029488.1:96. .300,
                     AY029489.1:118. .203,AY029490.1:1. .107,AY029491.1:34. .165,
                     AY029492.1:713. .843,AY029493.1:58. .310,AY029494.1:88. .305,
                     AY029495.1:75. .269,AY029496.1:102. .669,
                     AY029497.1:185. .288,AY029497.1:383. .509,
                     AY029498.1:99. .283,AY029499.1:88. .415,AY029500.1:145. .221,
                     AY029501.1:119. .244,AY029502.1:148. .348,
                     AY029503.1:177. .257,AY029503.1:337. .441,105. .292,
                     439. .1072)
                     /gene="ATP10C"
                     /codon_start=1
                     /product="aminophospholipid-transporting ATPase"
                     /protein_id="AAK33100.1"
```

```
                     /db_xref="GI:13878299"
                     /translation="MEREPAGTEEPGPPGRRRRREGRTRTVRSNLLPPPGAEDPAAGA
                     AKGERRRRRGCAQHLADNRLKTTKYTLLSFLPKNLFEQFHRPANVYFVFIALLNFVPA
                     VNAFQPGLALAPVLFILAITAFRDLWEDYSRHRSDHKINHLGCLVFSREEKKYVNRFW
                     KEIHVGDFVRLRCNEIFPADILLLSSSDPDGLCHIETANLDGETNLKRRQVVRGFSEL
                     VSEFNPLTFTSVIECEKPNNDLSRFRGCIIHDNGKKAGLYKENLLLRGCTLRNTDAVV
                     GIVIYAGHETKALLNNSGPRYKRSKLERQMNCDVLWCVLLLVCMSLFSAVGHGLWIWR
                     YQEKKSLFYVPKSDGSSLSPVTAAVYSFLTMIIVLQVLIPISLYVSIEIVKACQVYFI
                     NQDMQLYDEETDSQLQCRALNITEDLGQIQYIFSDKTGTLTENKMVFRRCTVSGVEYS
                     HDANAQRLARYQEADSEEEEVVPRGGSVSQRGSIGSHQSVRVVHRTQSTKSHRRTGSR
                     AEAKRASMLSKHTAFSSPMEKDITPDPKLLEKVSECDKSLAVARHQEHLLAHLSPELS
                     DVFDFFIALTICNTVVVTSPDQPRTKVRVRFELKSPVKTIEDFLRRFTPSCLTSGCSS
                     IGSLAANKSSHKLGSSFPSTPSSDGMLLRLEERLGQPTSAIASNGYSSQADNWASELA
                     QEQESERELRYEAESPDEAALVYAARAYNCVLVERLHDQVSVELPHLGRLTFELLHIL
                     GFDSVRKRMSVVIRHPLTDEINVYTKGADSVVMDLLQPCSSVDARGRHQKKIRSKTQN
                     YLNVYAAEGLRTLCIAKRVLSKEEYACWLQSHLEAESSLENSEELLFQSAIRLETNLH
                     LLGATGIEDRLQDGVPETISKLRQAGLQIWVLTGDKQETAVNIAYACKLLDHDEEVIT
                     LNATSQEACAALLDQCLCYVQSRGLQRAPEKTKGKVSMRFSSLCPPSTSTASGRRPSL
                     VIDGRSLAYALEKNLEDKFLFLAKQCRSVLCCRSTPLQKSMVVKLVRSKLKAMTLAIG
                     DGANDVSMIQVADVGVGISGQEGMQAVMASDFAVPKFRYLERLLILHGHWCYSRLANM
                     VLYFFYKNTMFVGLLFWFQFFCGFSASTMIDQWYLIFFNLLFSSLPPLVTGVLDRDVP
                     ANVLLTNPQLYKSGQNMEEYRPRTFWFNMADAAFQSLVCFSIPYLAYYDSNVDLFTWG
                     TPIVTIALLTFLLHLGIETKTWTWLNWITCGFSVLLFFTVALIYNASCATCYPPSNPY
                     WTMQALLGDPVFYLTCLMTPVAALLPRLFFRSLQGRVFPTQLQLARQLTRKSPRRCSA
                     PKETFAQGRLPKDSGTEHSSGRTVKTSVPLSQPSWHTQQPVCSLEASGEPSTVDMSMP
                     VREHTLLEGLSAPAPMSSAPGEAVLRSPGGCPEESKVRAASTGRVTPLSSLFSLPTFS
                     LLNWISSWSLVSRLGSVLQFSRTEQLADGQAGRGLPVQPHSGRSGLQGPDHRLLIGAS
                     SRRSQ"
     exon            105. .292
                     /gene="ATP10C"
                     /number=20
     exon            439. .1072
                     /gene="ATP10C"
                     /number=21
     3'UTR           1073. .3140
                     /gene="ATP10C"
BASE COUNT      789 a    784 c    789 g    778 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 3140;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 tcgcaagtcacagaa 22
           |||||||||||||||
Db    2244 TCGCAAGTCACAGAA 2258


RESULT  30
AX136409/c
LOCUS       AX136409      3520 bp    DNA             PAT       30-MAY-2001
DEFINITION  Sequence 331 from Patent EP1067182.
ACCESSION   AX136409
VERSION     AX136409.1  GI:14272813
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3520)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and
            Hayashi,K.
  TITLE     Secretory protein or membrane protein
  JOURNAL   Patent: EP 1067182-A 331 10-JAN-2001;
            Helix Research Institute (JP)
FEATURES             Location/Qualifiers
     source          1. .3520
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             40. .1878
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAC39845.1"
                     /db_xref="GI:14272814"
                     /translation="MAGHSMVLVEILDVNDNVPEVMVTSLSLPVQEDAQVGTVIALIS
                     VSDRDSGANGQVICSLTPHVPFKLVSTYKNYYSLVLDSALDRESVSAYELVVTARDGG
                     SPSLWATARVSVEVADVNDNAPAFAQPEYTVFVKENNPPGCHIFTVSAWDADAQKNAL
                     VSYSLVERRVGEHALSSYVSVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGS
                     NVTLQVFVLDENDNAPALLATPAGSAGGAVSELVPRSVGAGHVVAKVRAVDADSGYNA
                     WLSYELQPAAVGAHIPFHVGLYTGEISTTRILDEADAPRHRLLVLVKDHGEPALTSTA
                     TVLVSLVENGQAPKTSSRASVGAVDPEAALVDINVYLIIAICAVSSLLVLTLLLYTAL
                     RCSAPPTVSRCAPGKPTLVCSSAVGSWSYSQQRRQRVCSAESPPKTDLMAFSPSLQLS
                     REDCLNPPSEPRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSAT
                     PEPEAGEVSPPVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPT
                     NSQIDKSDFITFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT      871 a    896 c    901 g    852 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 6;  Length 3520;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db    3182 CAAGTCACAGAATTA 3168


RESULT  31
AX098224/c
LOCUS       AX098224      3748 bp    DNA             PAT       30-MAR-2001
DEFINITION  Sequence 136 from Patent WO0118542.
ACCESSION   AX098224
VERSION     AX098224.1  GI:13515348
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3748)
  AUTHORS   Lee,J., Thompsho,P. and Lillie,J.
  TITLE     Identification, assessment, prevention, and therapy of ovarian
            cancer
  JOURNAL   Patent: WO 0118542-A 136 15-MAR-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .3748
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     1110 a    825 c    761 g   1052 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 6;  Length 3748;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db    2627 CAAGTCACAGAATTA 2613


RESULT  32
D13644/c
LOCUS       D13644       4602 bp    mRNA            PRI       14-JUN-2001
DEFINITION  Homo sapiens mRNA for KIAA0019 protein, complete cds.
ACCESSION   D13644
VERSION     D13644.1  GI:1531551
KEYWORDS    KIAA0019.
SOURCE      Homo sapiens male myoblast cell_line:KG-1 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y.,
```

```
            Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
  TITLE     Prediction of the coding sequences of unidentified human genes. I.
            The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
            analysis of randomly sampled cDNA clones from human immature
            myeloid cell line KG-1
  JOURNAL   DNA Res. 1 (1), 27-35 (1994)
  MEDLINE   96051387
REFERENCE   2  (sites)
  AUTHORS   Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y.,
            Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
  TITLE     Prediction of the coding sequences of unidentified human genes. I.
            The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
            analysis of randomly sampled cDNA clones from human immature
            myeloid cell line KG-1 (supplement)
  JOURNAL   DNA Res. 1 (1), 47-56 (1994)
  MEDLINE   96051389
REFERENCE   3  (bases 1 to 4602)
  AUTHORS   Matoskova,B., Wong,W.T., Seki,N., Nagase,T., Nomura,N.,
            Robbins,K.C. and Di Fiore,P.P.
  TITLE     RN-tre identifies a family of tre-related proteins displaying a
            novel potential protein binding domain
  JOURNAL   Oncogene 12 (12), 2563-2571 (1996)
  MEDLINE   96293402
REFERENCE   4  (bases 1 to 4602)
  AUTHORS   Nomura,N., Miyajima,N., Kawarabayasi,Y. and Tabata,S.
  TITLE     Prediction of new human genes by entire sequencing of randomly
            sampled cDNA clones
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 4602)
  AUTHORS   Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1992) Nobuo Nomura, Kazusa DNA Research
            Institute, Gene Structure 1; 1532-3 Yana, Kisarazu, Chiba 292,
            Japan (E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp,
            Tel:0438-52-3930, Fax:0438-52-3931)
COMMENT     On Sep 10, 1996 this sequence version replaced gi:286016.
FEATURES             Location/Qualifiers
     source          1. .4602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10p13"
                     /sex="male"
                     /cell_line="KG-1"
                     /cell_type="myoblast"
     5'UTR           <1. .279
     gene            280. .2766
                     /gene="KIAA0019"
     CDS             280. .2766
                     /gene="KIAA0019"
                     /note="protein related N-ternimus of tre oncogene"
                     /codon_start=1
                     /product="KIAA0019 protein"
                     /protein_id="BAA02807.1"
                     /db_xref="GI:2104571"
                     /translation="MNSDQDVALKLAQERAE1VAKYDRGREGAEIEPWEDADYLVYKV
                     TDRFGFLHEEELPDHNVAVERQKHLEIERTTKWLKMLKGWEKYKNTEKFHRRIYKGIP
                     LQLRGEVWALLLEIPKMKEETRDLYSKLKHRARGCSPDIRQIDLLDVNRTFRDHIMFRD
                     RYGVKQQSLFHVLAAYSIYNTEVGYCQGMSQITALLLMYMNEEDAFWALVKLFSGPKH
                     AMHGFFVQGFPKLLRFQEHFEKILNKFLSKLKQHLDSQEIYTSFYTMKWFFQCFLDRT
                     PFTLNLRIWDIYIFEGERVLTAMSYTILKLHKKHLMKLSMEELVEFFQETLAKDFFFE
                     DDFVIEQLQISMTELKRAKLDLPEPGKEDEYPKKPLGQLPPELQSWGVHHLSNGQRSV
                     GRPSPLASGRRESGAPHRRHEHSPHPQSRTGTPERAQPPRRKSVEEESKKLKDEADFQ
                     RKLPSGPQDSSRQYNHAAANQNSNATSNIRKEFVPKWNKPSDVSATERTAKYTMEGKG
                     RAAHPALAVTVPGPAEVRVSNVRPKMKALDAEDGKRGSTASQYDNVPGPELDSGASVE
                     EALERAYSQSPRHALYPPSPRKHAEPSSSPSKVSNKFTFKVQPPSHARYPSQLDGEAR
                     GLAHPPSYSNPPVYHGNSPKHFPTANSSFASPQFSPGTQLNPSRRPHGSTLSVSASPE
                     KSYSRPSPLVLPSSRIEVLPVDTGAGGYSGNSGSPKNGKLIIPPVDYLPDNRTWSEVS
                     YTYRPETQGQSWTRDASRGNLPKYSAFQLAPFQDHGLPAVSVDSPVRYKASPAAEDAS
                     PSGYPYSGPPPPAYHYRNRDGLSIQESVLL"
     3'UTR           2767. .4602
BASE COUNT     1340 a   1041 c    999 g   1222 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 9;  Length 4602;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4462 CAAGTCACAGAATTA 4448


RESULT  33
AF039582
LOCUS       AF039582     4636 bp    DNA             BCT       05-DEC-2000
DEFINITION  Enterobacter aerogenes putative integrase (int(p)) gene, partial
            cds; and EaeI methyltransferase alpha subunit (eaeIM-a), EaeI
            methyltransferase beta subunit (eaeIM-b), and EaeI restriction
            endonuclease (eaeIR) genes, complete cds.
ACCESSION   AF039582
VERSION     AF039582.1  GI:2760953
KEYWORDS    .
SOURCE      Enterobacter aerogenes.
  ORGANISM  Enterobacter aerogenes
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Enterobacter.
REFERENCE   1  (bases 1 to 4636)
  AUTHORS   Lee,K.F., Shaw,P.C., Picone,S.J., Wilson,G.G. and Lunnen,K.D.
  TITLE     Sequence comparison of the EcoHK31I and EaeI
            restriction-modification systems suggests an intergenic transfer of
            genetic material
  JOURNAL   Biol. Chem. 379 (4-5), 437-441 (1998)
  MEDLINE   98290299
   PUBMED   9628335
REFERENCE   2  (bases 1 to 4636)
  AUTHORS   Lunnen,K.D., Picone,S.J. and Wilson,G.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-1997) Research, New England Biolabs, 32 Tozer
            Road, Beverly, MA 01915, USA
FEATURES             Location/Qualifiers
     source          1. .4636
                     /organism="Enterobacter aerogenes"
                     /db_xref="taxon:548"
     gene            <1. .1035
                     /gene="int(p)"
     CDS             <1. .1035
                     /gene="int(p)"
                     /note="similar to phage P4 integrase SP:P39347"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integrase"
                     /protein_id="AAB95339.1"
                     /db_xref="GI:2760957"
                     /translation="KLVAAGIDPREHKRAVKEEQTKGAITFESVAREWHAANKKWTEE
                     HSRRVLKSLEDNLFPAIGKRSIDSFSTRDLLVPIKVVEATGRLEVASRLQQRTNAIMR
                     YAVQSGLIDYNPAQEMAGAVASSNRVHRPALELKRLPELLQRIDSYSGRPLTRLAVEL
                     TLLIFIRSSELRFARWPEIDFERAMWTIPAEREAIEGVKHSQRGSKMRTPHLVPLSRQ
                     ALEILKQVHKFSGERDFVFVGDHNPRKPMSENTVNKALRVMGYDTKVEVCGHGFRTMA
                     CSSLIESGLWSRDAVERQMSHMERNSVRAAYIHKAEHLDERKLMLQWWADFLDANRQK
                     GISPFDYAKA"
     gene            1228. .2157
                     /gene="eaeIM-a"
     CDS             1228. .2157
                     /gene="eaeIM-a"
                     /function="converts YGGCCR to YGGmCCR"
                     /note="M.EaeIa"
                     /codon_start=1
                     /transl_table=11
                     /product="EaeI methyltransferase alpha subunit"
                     /protein_id="AAB95336.1"
                     /db_xref="GI:2760954"
                     /translation="MKKKPLKQYKVSSFFAGIGGFDLGLEKAGMEVVFQCEINKFCQK
                     VLRKNWPNVPLHTDITQLNADEIPESSVWCGGFPCQDVSSANQGKRKGLEGARSGLFY
                     TYAKLIEERKPEWLVIENVPGLLNSHNGQDFKVVIDTLVELGYGVSWRVLDAKYFGTP
```

```
                     QRRRRVYIVASLGDMRSARVLFEPGATRIVDKQGQGERTNASGASDTGLSKADIYSIQ
                     HASIGRKATAGPQAKGYRNDGETYTLDSRGSSDAVCAPDAAFRVRASAGISRELDSNR
                     FRAVGNAVAVPIVEWIGKRIVLVDQQMLNEALS"
     gene            1538. .2068
                     /gene="eaeIM-b"
     CDS             1538. .2068
                     /gene="eaeIM-b"
                     /note="M.EaeIb; translated from 3'-half of the eaeIM-a
                     gene in an alternative reading frame"
                     /codon_start=1
                     /transl_table=11
                     /product="EaeI methyltransferase beta subunit"
                     /protein_id="AAB95337.1"
                     /db_xref="GI:2760955"
                     /translation="MQNSSKKENLNGLLLKMFPDCSTATMDKTSKLSSIRWSNSGMAF
                     RGEYWMQNTLEHPSVEEECTLSQVLETCAPLESFLNPEQLESLINRAKERGQMLPEPL
                     IQAYQKQISILSSMQVLDEKQPQDLKQKDTGTTEKPTHLTQEEVQMLYVRRMLPSEYE
                     RLQGFPENWTLIDSEQ"
     gene            complement(2138. .3085)
                     /gene="eaeIR"
     CDS             complement(2138. .3085)
                     /gene="eaeIR"
                     /function="cleaves between Y and G in YGGCCR"
                     /note="R.EaeI; unable to cleave methylated ds sequence
                     YGGmCCR"
                     /codon_start=1
                     /transl_table=11
                     /product="EaeI restriction endonuclease"
                     /protein_id="AAB95338.1"
                     /db_xref="GI:2760956"
                     /translation="MAVSIAEYLGQRTDVDQQIVPVSKGTLAQCPFMDRTCDKVSKNN
                     HPVCSVRKPDSTIWIVCEHRLCSTRQKKTVISGERTQQVDNVLVEHQRDILRKVAKLI
                     YQDPELQDSEIGVKREVNIPLPDSDNSYHADYVMRNFSGRGRVDEVLLEMQGGGETSS
                     TGAITRHIAAWADLEFPTNEILRRPVAANTIETNAWRRQQEQFLVKGNVVDQTGGKIV
                     FAVGSLLYDYLHRRFRNANLRDLKAHNWTLCILAFKEDTSDEPQPGPIPLIIDDSKTL
                     FTNYSTFVRFLTDQGAPRPELFEGSFLRLDNSTVTIMPR"
BASE COUNT     1362 a    893 c   1074 g   1307 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 1;  Length 4636;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 acattgtcgcaagtc 16
           ||||||||||||||| 
Db    1727 ACATTGTCGCAAGTC 1741


RESULT  34
AF152313/c
LOCUS       AF152313     5218 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 5 (PCDH-alpha5) mRNA, complete
            cds.
ACCESSION   AF152313
VERSION     AF152313.1  GI:5456913
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5218)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5218)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5218)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5218
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5218
                     /gene="PCDH-alpha5"
     CDS             1. .2811
                     /gene="PCDH-alpha5"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 5"
                     /protein_id="AAD43707.1"
                     /db_xref="GI:5456914"
                     /translation="MVYSRRGSLGSRLLLLWLLLAYWKAGSGQLHYSIPEEAKHGTFV
                     GRIAQDLGLELAELVPRLFRVASKGRGDLLEVNLQNGILFVNSRIDREELCRRRAECS
                     IHLEVIVDRPLQVFHVEVAVKDINDNPPRFSRQEQRLFILESRMPDSRFPLEGASDLD
                     IGANAQLRYRLNPNEYFDLDVKTNEEETNFLELVLRKSLDREETQEHRLLVIATDGGK
                     PELTGTVQLLINVLDANDNAPEFDKSIYNVRLLENAPSGTLVIKLNASDADEGINKEI
                     VYFFSNLVLDDVKSKFIINSNTGEIKVNGELDYEDYNSYEINIDAMDKSTFPLSGHCK
                     VVVKLLDVNDNTPEMAITTLFLPVKEDAPLSTVIALISVSDRDSGANGQVTCSLMPHV
                     PFKLVSTFKNYYSLVLDSALDRESVSVYELVVTARDGGSPSLWATASVSVEVADVNDN
                     APAFAQPQYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRVGERPLSSYVS
                     VHAESGKVYALQPLDHEEVELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALLV
                     PRVGGTGGAVSELVPRSVGAGHVVAKVRAVDPDSGYNAWLSYELQPAPGSARIPFRVG
                     LYTGEISTTRSLDETEAPRHRLLVLVKDHGEPPLTATATVLVSLVESGQAPKASSRAS
                     AGAVGPEAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSAQPTEAVCTRGKPTLLC
                     SSAVGSWSYSQQRRQRVCSGEAPPKTDLMAFSPSLPQGPTSTDNPRQPNPDWRYSASL
                     RAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPVGAGVNSNSWTFK
                     YGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITFGKKEETKKKKKK
                     KKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1385 a   1207 c   1281 g   1345 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 5218;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||| 
Db    4115 CAAGTCACAGAATTA 4101


RESULT  35
AF152315/c
LOCUS       AF152315     5221 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 7 (PCDH-alpha7) mRNA, complete
            cds.
ACCESSION   AF152315
VERSION     AF152315.1  GI:5456917
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5221)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5221)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
```

```
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5221)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5221
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5221
                     /gene="PCDH-alpha7"
     CDS             1. .2814
                     /gene="PCDH-alpha7"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 7"
                     /protein_id="AAD43709.1"
                     /db_xref="GI:5456918"
                     /translation="MVCPNGYDPGGRHLLLFIITLAAWEAGRGQLHYSVPEEAKHGNF
                     VGRIAQDLGLELAELVPRLFRAVCKFRGDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVERPLQVFHVDVEVKDINDNPPVFPATQRNLFIAESRPLDSRFPLEGASDA
                     DIGENALLTYRLSPNEYFFLDVPTSNQQVKPLGLVLRKLLDREETPELHLLLTATDGG
                     KPELTGTVQLLITVLDNNDNAPVFDRTLYTVKLPENVSIGTLVIHPNASDLDEGLNGD
                     IIYSFSSDVSPDIKSKFHMDPLSGAITVIGHMDFEESRAHKIPVEAVDKGFPPLAGHC
                     TVLVEVVDVNDNAPQLTLTSLSLPIPEDAQPGTVITLISVFDRDFGVNGQVTCSLTPR
                     VPFKLVSTFKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATASVSVEVADVNN
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSAGDADAQKNALVSYSLVELRVGERALSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     APRVGGTGGAVRELVPRSVGAGHVVAKVRAVDADSGYNAWLSYELQPVAAGASIPFRV
                     GLYTGEISTTRALDETDAPRHRLLVLVKDHGEPSLTATATVLVSLVESGQAPKASSRA
                     SLGIAGPETELVDVNVYLIIAICAVSSLLVLTLLLYTALRCSAPSSEGACSLVKPTLV
                     CSSAVGSWSFSQQRRQRVCSGEGPPKTDLMAFSPSLPQGPSSTDNPRQPNPDWRYSAS
                     LRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPVGAGVNSNSWTF
                     KYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITFGKKEETKKKKK
                     KKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1333 a    1242 c    1295 g    1351 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 5221;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||| |
Db    4118 CAAGTCACAGAATTA 4104


RESULT  36
AF152308/c
LOCUS       AF152308     5233 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 12 (PCDH-alpha12) mRNA, complete
            cds.
ACCESSION   AF152308
VERSION     AF152308.1  GI:5456901
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5233)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5233)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5233)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5233
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5233
                     /gene="PCDH-alpha12"
     CDS             1. .2826
                     /gene="PCDH-alpha12"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 12"
                     /protein_id="AAD43702.1"
                     /db_xref="GI:5456902"
                     /translation="MVIIGPRGPGSQRLLLSLLLLAAWEVGSGQLHYSVYEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRVASKRHGDLLEVNLQNGILFVNSRIDREKLCGRSAEC
                     SIHLEVIVDRPLQVFHVDVEVKDINDNPPVFREREQKVPVSESAPLDSHFPLEGASDA
                     DIGVNSLLTYALSLNENFELKIKTKKDKSILPELVLRKLLDREQTPKLNLLLMVIDGG
                     KPELTGSVQIQITVLDVNDNGPAFDKPSYKVVLSENVQNDTRVIQLNASDPDEGLNGE
                     ISYGIKMILPVSEKCMFSINPDTGEIRIYGELDFEENNAYEIQVNAIDKGIPSMAGHS
                     MVLVEVLDVNDNVPEVMVTSLSLPVQEDAQVGTVIALISVSDRDSGANGQVICSLTPH
                     VPFKLVSTYKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATARVSVEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSAWDADAQKNALVSYSLVERRVGEHALSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     ATPAGSAGGAVSELVPRSVGAGHVVAKVRAVDADSGYNAWLSYELQPAAVGAHIPFHV
                     GLYTGEISTTRILDEADAPRHRLLVLVKDHGEPALTSTATVLVSLVENGQAPKTSSRA
                     SVGAVDPEAALVDINVYLIIAICAVSSLLVLTLLLYTALRCSAPPTVSRCAPGKPTLV
                     CSSAVGSWSYSQQRRQRVCSAESPPKTDLMAFSPSLQLSREDCLNPPSEPRQPNPDWR
                     YSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPVGAGVNSN
                     SWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITFGKKEETK
                     KKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1364 a    1213 c    1305 g    1351 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 5233;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||| |
Db    4130 CAAGTCACAGAATTA 4116


RESULT  37
AF152312/c
LOCUS       AF152312     5251 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 4 (PCDH-alpha4) mRNA, complete
            cds.
ACCESSION   AF152312
VERSION     AF152312.1  GI:5456911
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5251)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
```

```
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5251)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5251)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5251
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5251
                     /gene="PCDH-alpha4"
     CDS             1. .2844
                     /gene="PCDH-alpha4"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 4"
                     /protein_id="AAD43706.1"
                     /db_xref="GI:5456912"
                     /translation="MEFSWGSGQESRRLLLLLLLLAAWEAGNGQLHYSVSEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRVASKGRGGLLEVNLQNGILFVNSRIDREELCRRSAEC
                     SIHLEVIVDRPLQVFHVDVEVRDINDNPPVFPATQKNLSIAESRPLDSRFPLEGASDA
                     DIGENALLTYRLSPNEYFSLEKPPDDELVKGLGLILRKSLDREEAPEIFLVLTATDGG
                     KPELTGTVQLLITVLDANDNAPAFDRTIYKVRLLENVPNGTLVIKLNASDLDEGLNGD
                     IVYSFSNDISPNVKSKFHIDPITGQITVKGYIDFEESKSYEIIVEGIDKGQLPLSGHC
                     RVIVEVEDNNDNVPDLEFKSLSLPIREDAPLGTVIALISVSDKDMGVNGLVTCSLTSH
                     VPFKLVSTFKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATASVSVEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSAWDADAQENALVSYSLVERRVGERALSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVTARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     APRAGGTGGAVSELVPWSVGVGHVVAKVRAVDADSGYNAWLSYELQPGTGGARIPFRV
                     GLYTGEISTTRALDETDAPRHRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRA
                     LVGAVGPDAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSALPTEGACAPGKPTLV
                     CSSAVGSWSYSQQRRPRVCSGEGPPKTDLMAFSPSLPDSRDREDQLQTTEESFAKPRQ
                     PNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPVG
                     AGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITFG
                     KKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1360 a   1229 c   1304 g   1358 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 5251;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4148 CAAGTCACAGAATTA 4134


RESULT  38
AF152306/c
LOCUS       AF152306     5254 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 10 (PCDH-alpha10) mRNA, complete
            cds.
ACCESSION   AF152306
VERSION     AF152306.1  GI:5456897
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5254
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5254
                     /gene="PCDH-alpha10"
     CDS             1. .2847
                     /gene="PCDH-alpha10"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 10"
                     /protein_id="AAD43700.1"
                     /db_xref="GI:5456898"
                     /translation="MVSRCSCLGVQCLLLSLLLLAAWEVGSGQLHYSVYEEARHGTFV
                     GRIAQDLGLELAELVQRLFRVASKRHGDLLEVNLQNGILFVNSRIDREELCGRSVECS
                     IHLEVIVDRPLQVFHVDVEVKDINDNPPRFSVTEQKLSIPESRLLDSRFPLEGASDAD
                     VGENALLTYKLSPNEYFVLDIINKKDKDKFPVLVLRKLLDREENPQLKLLLTATDGGK
                     PEFTGSVSLLILVLDANDNAPIFDRPVYEVKMYENQVNQTLVIRLNASDSDEGINKEM
                     MYSFSSLVPPTIRRKFWINERTGEIKVNDAIDFEDSNTYEIHVDVTDKGNPPMVGHCT
                     VLVELLDENDNSPEVIVTSLSLPVKEDAQVGTVIALISVSDHDSGANGQVTCSLTPHV
                     PFKLVSTYKNYYSLVLDSALDRERVSAYELVVTARDGGSPPLWATASVSVEVADVNDN
                     APAFAQSEYTVFVKENNPPGCHIFTVSAWDADAQENALVSYSLVERRLGERSLSSYVS
                     VHAESGKVYALQPLDHEELELLQFQVSARDGGVPPLGSNLTLQVFVLDENDNAPALLA
                     SPAGSAGGAVSELVLRSVVAGHVVAKVRAVDADSGYNAWLSYELQSAAVGARIPFRVG
                     LYTGEISTTRALDETDSPRQRLLVLVKDHGEPSLTATATVLVSLVEGSQAPKASSRAS
                     VGVAPEVALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSAAPTEGACGPVKPTLVCS
                     SAVGSWSYSQQRRQRVCSGEGLPKADLMAFSPSLPPCPMVDVDGEDQSIGGDHSRKPR
                     QPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPV
                     GAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITF
                     GKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1358 a   1210 c   1317 g   1369 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 5254;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4151 CAAGTCACAGAATTA 4137


RESULT  39
AF152310/c
LOCUS       AF152310     5254 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 2 (PCDH-alpha2) mRNA, complete
            cds.
ACCESSION   AF152310
VERSION     AF152310.1  GI:5456905
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
```

```
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5254)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5254
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5254
                     /gene="PCDH-alpha2"
     CDS             1. .2847
                     /gene="PCDH-alpha2"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 2"
                     /protein_id="AAD43704.1"
                     /db_xref="GI:5456906"
                     /translation="MASSIRRGRGAWTRLLSLLLLAAWEVGSGQLRYSVPEEAKHGTF
                     VGRIAQDLGLELEELVPRLFRVASKRHGDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHVEVlVDRPLQVFHVEVEVKDINDNPPIFPMTVKTIRFPESRLLDSRFPLEGASDA
                     DIGVNALLSYKLSSSEFFFLDIQANDELSESLSLVLGKSLDREETAEVNLLLVATDGG
                     KPELTGTVQILIKVLDVNDNEPTFAQSVYKVKLLENTANGTLVVKLNASDADEGPNSE
                     IVYSLGSDVSSTIQTKFTIDPISGEIRTKGKLDYEEAKSYEIQVTATDKGTPSMSGHC
                     KISLKLVDINDNTPEVSITSLSLPISENASLGTVIALITVSDRDSGTNGHVTCSLTPH
                     VPFKLVSTFKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATTSVSIEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSAWDADAQENALVSYSLVERRVGERALSSYV
                     SVHAESGKVYALQPLDHEEVELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     APRAGTAAGAVSELVPWSVGAGHVVAKVRAVDADSGYNAWLSYELQLGTGSARIPFRV
                     GLYTGEISTTRALDEADSPRHRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRA
                     WVGAAGSEATLVDVNVYLIIAICAVSSLLVLTVLLYTALRCSVPPTEGARAPGKPTLV
                     CSSAVGSWSYSQQRRQRVCSGEDPPKTDLMAFSPSLSQGPDSAEEKQLSESEYVGKPR
                     QPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPPV
                     GAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFITF
                     GKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1375 a   1234 c   1300 g   1345 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 5254;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4151 CAAGTCACAGAATTA 4137


RESULT  40
AF152307/c
LOCUS       AF152307     5257 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 11 (PCDH-alpha11) mRNA, complete
            cds.
ACCESSION   AF152307
VERSION     AF152307.1  GI:5456899
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5257)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5257)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5257)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5257
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5257
                     /gene="PCDH-alpha11"
     CDS             1. .2850
                     /gene="PCDH-alpha11"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 11"
                     /protein_id="AAD43701.1"
                     /db_xref="GI:5456900"
                     /translation="MFGFQRRGLGTPRLQLWLLLLEFWEVGSGQLHYSVSEEAKHGTF
                     VGRIAQDLGLELAELVQRLFRVASKTHGDLLEVNLQNGILFVNSRIDREELCGQSAEC
                     SIHLEVIVDRPLQVFHVNVEVKDINDNPPVFSLREQKLLIAESKQSDSRFPLEGASDA
                     DIEENALLTYRLSKNEYFSLDSPTNGKQIKRLSLILKKSLDREKTPELNLLLTATDGG
                     KPELTGTVRLLVQVLDVNDNDPEFDKSEYKVSLMENAAKETLVLKLNATDRDEGVNGE
                     VTYSLMSIKPNGRHLFTLDQNNGEVRVNGTLDYEENKFYKIEVQATDKGTPPMAGHCT
                     VWVEILDTNDNSPEVAVTSLSLPVREDAQPSTVIALISVSDRDSGVNGQVTCSLTPHV
                     PFKLVSTFKNYYSLVLDSALDRENVWAYELVVTARDGGSPSLWATARVSVEVADVNDN
                     APAFAQPEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRLGDRALSSYVS
                     VHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLSSNVTLQVFVLDENDNAPALLA
                     TQAGSAGGAVNKLVPRSVGAGHVVAKVRAVDADSGYNAWLSYELQPAAGGSRIPFRVG
                     LYTGEISTTRALDEADSPRHRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRTL
                     AGAASPEAALVDVNVYLIIAICVVSSLLVLTLLLYTALWWSATPTEGACAPGKPTLVC
                     SRAVGSWSYSQQRRQRVCSEEGPPKTDLMAFSPSLPLGLNKEEEGERQEPGSNHPGQP
                     RQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSPP
                     VGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFIT
                     FGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1393 a   1224 c   1317 g   1323 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 5257;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4154 CAAGTCACAGAATTA 4140


RESULT  41
AF152305/c
LOCUS       AF152305     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 1 (PCDH-alpha1) mRNA, complete
            cds.
ACCESSION   AF152305
VERSION     AF152305.1  GI:5456895
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha1"
     CDS             1. .2853
                     /gene="PCDH-alpha1"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 1"
                     /protein_id="AAD43699.1"
                     /db_xref="GI:5456896"
                     /translation="MVFSRRGGLGARDLLLWLLLLAAWEVGSGQLHYSIPEEAKHGTF
                     VGRVAQDLGLELAELVPRLFRVASKTHRDLLEVNLQNGILFVNSRIDREELCQWSAEC
                     SIHLELIADRPLQVFHVEVKVKDINDNPPVFRGREQIIFIPESRLLNSRFPIEGAADA
                     DIGANALLTYTLSPSDYFSLDVEASDELSKSLWLELRKYLDREETPELHLLLTATDGG
                     KPELQGTVELLITVLDVNDNAPLFDQAVYRVHLLETTANGTLVTTLNASDADEGVNGE
                     VVFSFDSGISRDIQEKFKVDSSSGEIRLIDKLDYEETKSYEIQVKAVDKGSPPMSNHC
                     KVLVKVLDVNDNAPELAVTSLYLPIREDAPLSTVIALITVSDRDSGANGQVTCSLMPH
                     VPFKLVSTFKNYYSLVLDSALDRESLSVYELVVTARDGGSPSLWATARVSVEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRVGERALSNYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     APRVGGTIGAVSELVPRLVGAGHVVAKVRAVDADSGYNAWLSYELQPAAGGARIPFRV
                     GLYTGEISTTRVLDEADLSRYRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRA
                     SVGVAGPEAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSVPPTEGAYVPGKPTLV
                     CSSALGSWSNSQQRRQRVCSSEGPPKTDLMAFSPGLSPSLNTSERNEQPEANLDLSGN
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRACPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1365 a   1217 c   1306 g   1372 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 5260;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4157 CAAGTCACAGAATTA 4143


RESULT  42
AF152309/c
LOCUS       AF152309     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 13 (PCDH-alpha13) mRNA, complete
            cds.
ACCESSION   AF152309
VERSION     AF152309.1  GI:5456903
KEYWORDS    .
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha13"
     CDS             1. .2853
                     /gene="PCDH-alpha13"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 13"
                     /protein_id="AAD43703.1"
                     /db_xref="GI:5456904"
                     /translation="MLSSWQGGPRPRQLLLWLLILAAWETGSGQLHYSVPEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRVASKRHGDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVDRPLQVFHVEVKVRDINDNPPIFPESKKRIIIAESRPPETRFPLDGASDA
                     DIGVNSALTYRLDPNDYFTLDAQNSLEQMSSLSLVLRKTLDREEIQEHSLLLTASDGG
                     KPELTGTVQLLITILDVNDNAPEFYQSVYKVTVLENAFNGTLVIKLNATDPDDGTNGD
                     IVYSFRRPVWPAVVYAFTINPNNGEIRTKGKLDFEEKKLYEISVEAVDKGNIPMAGHC
                     TLLVEVLDVNDNAPEVTITSLSLPIREDTQPSAIIALISVSDRDSGSNGQVTCTLTPH
                     VPFKLVSTYKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATASVSVGVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSAQDADAQENALVSYSLVERRVGERALSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDSGVPPLGSNVTLQVFVLDENDNAPALL
                     TPGAGSAGGTVSELMPRSVGAGHVVAKVRAVDADSGYNAWLSYELQLAAVGARIPFRV
                     GLYTGEISTTRPLDEVDAPHHRLLVLVKDHGEPALTATATVLLSLVESGQAPQASSRA
                     SAGAVGPEAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSAPPTEGACAPGKPTLV
                     CSSAAGSWSYSQQRRPRVCSGEGPHKTDLMAFSPSLPPCLGSAEGTGQREEDSECLKE
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1376 a   1234 c   1315 g   1335 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 5260;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4157 CAAGTCACAGAATTA 4143


RESULT  43
AF152311/c
LOCUS       AF152311     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 3 (PCDH-alpha3) mRNA, complete
            cds.
ACCESSION   AF152311
```

```
VERSION     AF152311.1  GI:5456909
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha3"
     CDS             1. .2853
                     /gene="PCDH-alpha3"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 3"
                     /protein_id="AAD43705.1"
                     /db_xref="GI:5456910"
                     /translation="MLFSWREDPGAQCLLLSLLLLAASEVGSGQLHYSVSEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRVASKRHGDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVDRPLQVFHVEVEVKDINDNAPVFPMAVKNLFISESRQPGSRFSLEGASDA
                     DIGTNSLLTYSLDSTEYFTLDVKRNDEEIKSLGLVLKKNLNREDTPKHYLLITAIDGG
                     KPELTGTTQLKITVLDVNDNAPAFERTIYKVRLLENAPNGTLVVTVNATDLDEGVNKD
                     IAYSFNTDMSADILSKFHLDPVNGQISVKGNIDFEESKSYEIQVEATDKGNPPMSDHC
                     TVLLEIVDINDNVPELVIQSLSLPVLEDSPLSTVIALISVSDRDSGVNGQVTCSLTPH
                     VPFKLVSTFKNYYSLVLDSPLDRESVSAYELVVTARDGGSPSLWATASVSVEVADVND
                     NAPAFSQSEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRVGERALSSYV
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     MPRVGGIGGAVSELVPRSVGAGHVVAKVRAVDADSGYNAWLSYELQPGTGGARIPFRV
                     GLYTGEISTTRALDEVDAPRHRLLVLVKDHGEPSLTATATVLVSLVESGQAPKASSQA
                     SAGATGPEAALVDVNVYLIVAICAVSSLLVLTLLLYTALRCSAPPTEGDCGPGKPTLV
                     CSSAVGSWSYSQQRQQRVCSGEGLPKTDLMAFSPSLPPCPISRDREEKQDVDVDLSAK
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1379 a   1225 c   1293 g   1363 t
ORIGIN


  Query Match           60.0%;  Score 15;  DB 9;  Length 5260;
  Best Local Similarity  100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4157 CAAGTCACAGAATTA 4143


RESULT  44
AF152314/c
LOCUS       AF152314     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 6 (PCDH-alpha6) mRNA, complete
            cds.
ACCESSION   AF152314
VERSION     AF152314.1  GI:5456915
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha6"
     CDS             1. .2853
                     /gene="PCDH-alpha6"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 6"
                     /protein_id="AAD43708.1"
                     /db_xref="GI:5456916"
                     /translation="MVFTPEDRLGKQCLLLPLLLLAAWKVGSGQLHYSVPEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRMASKDREDLLEVNLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVDRPLQVFHVDVEVRDINDNPPLFPVEEQRVLIYESRLPDSVFPLEGASDA
                     DVGSNSILTYKLSSSEYFGLDVKINSDDNKQIGLLLKKSLDREEAPAHNLFLTATDGG
                     KPELTGTVQLLVTVLDVNDNAPTFEQSEYEVRIFENADNGTTVIRLNASDRDEGANGA
                     ISYSFNSLVAAMVIDHFSIDRNTGEIVIRGNLDFEQENLYKILIDATDKGHPPMAGHC
                     TVLVRILDKNDNVPEIALTSLSLPVREDAQFGTVIALISVNDLDSGANGQVNCSLTPH
                     VPFKLVSTFKNYYSLVLDSALDRESVSAYELVVTARDGGSPSLWATASLSVEVADMND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRVGERALSSYI
                     SVHAESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     APRVGGTGGAVSELVPRSLGAGQVVAKVRAVDADSGYNAWLSYELQPPASSARFPFRV
                     GLYTGEISTTRVLDEADSPRHRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRA
                     SVGAAGPEAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSAPPTEGACTADKPTLV
                     CSSAVGSWSYSQQRRQRVCSGEGPPKMDLMAFSPSLSPCPIMMGKAENQDLNEDHDAK
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDFI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1366 a   1230 c   1315 g   1349 t
ORIGIN


  Query Match           60.0%;  Score 15;  DB 9;  Length 5260;
  Best Local Similarity  100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    4157 CAAGTCACAGAATTA 4143


RESULT  45
AF152316/c
```

```
LOCUS       AF152316     5260 bp    mRNA            PRI       22-JUL-1999
DEFINITION  Homo sapiens protocadherin alpha 8 (PCDH-alpha8) mRNA, complete
            cds.
ACCESSION   AF152316
VERSION     AF152316.1  GI:5456919
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     A striking organization of a large family of human neural
            cadherin-like cell adhesion genes
  JOURNAL   Cell 97 (6), 779-790 (1999)
  MEDLINE   99308636
   PUBMED   10380929
REFERENCE   2  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Human protocadherin genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5260)
  AUTHORS   Wu,Q. and Maniatis,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAY-1999) Department of Molecular and Cellular
            Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA
            02138, USA
FEATURES             Location/Qualifiers
     source          1. .5260
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5q31"
                     /tissue_type="brain"
     gene            1. .5260
                     /gene="PCDH-alpha8"
     CDS             1. .2853
                     /gene="PCDH-alpha8"
                     /function="cell-cell adhesion"
                     /codon_start=1
                     /product="protocadherin alpha 8"
                     /protein_id="AAD43710.1"
                     /db_xref="GI:5456920"
                     /translation="MDYHWRGELGSWRLLLLLLLLAAWKVGSGQLHYSVPEEAKHGTF
                     VGRIAQDLGLELAELVPRLFRVASKRHRDLLEVSLQNGILFVNSRIDREELCGRSAEC
                     SIHLEVIVDRPLQVFHVDVEVKDVNDNPPVFRVKDQKLFVSESRMPDSRFPLEGASDA
                     DVGANSVLTYRLSSHDYFMLDVNSKNDENKLVELVLRKSLDREDAPAHHLFLTATDGG
                     KPELTGTVQLLVTVLDVNDNAPTFEQSEYEVRIFENADNGTTVIKLNASDPDEGANGA
                     ISYSFNSLVETMVIDHFSIDRNTGEIVIRGNLDFEQENLYKILIDATDKGHPPMAGHC
                     TVLVRILDKNDNVPEIALTSLSLPVREDAQFGTVIALISVNDLDSGANGQVTCSLMPH
                     VPFKLVSTFKNYYSLVLDSALDRERVSAYELVVTARDGGSPSLWATASLSVEVADVND
                     NAPAFAQPEYTVFVKENNPPGCHIFTVSARDADAQENALVSYSLVERRVGERSLSSYI
                     SVHTESGKVYALQPLDHEELELLQFQVSARDAGVPPLGSNVTLQVFVLDENDNAPALL
                     EPRVGGTGGAASKLVPRSVGAGHVVAKVRAVDADSGYNAWLSYELQPAASSPRIPFRV
                     GLYTGEISTTRVLDEADSPRHRLLVLVKDHGEPALTATATVLVSLVESGQAPKASSRQ
                     SAGVLGPEAALVDVNVYLIIAICAVSSLLVLTLLLYTALRCSALPTEGGCRAGKPTLV
                     CSSAVGSWSYSQQQPQRVCSGEGPPKTDLMAFSPCLPPDLGSVDVGEEQDLNVDHGLK
                     PRQPNPDWRYSASLRAGMHSSVHLEEAGILRAGPGGPDQQWPTVSSATPEPEAGEVSP
                     PVGAGVNSNSWTFKYGPGNPKQSGPGELPDKFIIPGSPAIISIRQEPTNSQIDKSDPI
                     TFGKKEETKKKKKKKKGNKTQEKKEKGNSTTDNSDQ"
BASE COUNT     1357 a   1236 c   1310 g   1357 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 9;  Length 5260;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||| 
Db    4157 CAAGTCACAGAATTA 4143
```

```
Search completed: December 26, 2001, 12:50:03
Job time: 8725 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:20:08 ; Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196F-1
Perfect score: 25
Sequence: 1 gacattgtcgcaagtcacagaatta 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 33925

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	839	22	AAI62941	Human genomic DNA
	2	25	100.0	1041	20	AAX01303	E. coli biotin syn
	3	25	100.0	1084	10	AAN91329	E.coli Bio B gene.
	4	25	100.0	1121	7	AAN60496	Sequence encoding
	5	25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
c	6	17	68.0	17	14	AAQ56307	BioB DNA primer.
	7	15	60.0	561	22	AAF94235	Primer specific fo
	8	15	60.0	1771	20	AAZ30602	Mouse integrin alp
	9	15	60.0	1771	22	AAS10264	Mouse integrin alp
c	10	15	60.0	3520	22	AAF93909	Human cDNA encodin
c	11	15	60.0	3520	22	AAF29361	Amyloid-beta prote
c	12	15	60.0	3748	22	AAF98729	Human late stage o
	13	15	60.0	6087	22	AAS22683	Human cDNA encodin
	14	15	60.0	6116	22	AAS22447	Human cDNA encodin
c	15	15	60.0	6567	11	AAQ03324	Eimeria tenella ge
	16	14	56.0	523	22	AAH12853	Human cDNA clone (
c	17	14	56.0	1050	9	AAN80927	Lividans Exported
c	18	14	56.0	1465	22	AAH14046	Human cDNA sequenc
c	19	14	56.0	1677	17	AAT31932	Retinoid X recepto
c	20	14	56.0	1787	17	AAT31928	Retinoid X recepto
c	21	14	56.0	1890	20	AAZ52940	Human prostate tum
	22	14	56.0	1919	22	AAI60385	Human polynucleoti
c	23	14	56.0	2022	22	AAH15029	Human cDNA sequenc
c	24	14	56.0	2102	22	AAI58599	Human polynucleoti
c	25	14	56.0	2220	22	AAD12786	Human nuclear horm
c	26	14	56.0	2233	19	AAV10120	Human retinoid rec
c	27	14	56.0	2585	22	AAF77541	Caenorhabditis bri
c	28	14	56.0	3083	12	AAQ11851	Glutamate receptor
c	29	14	56.0	5662	22	AAI61089	Human polynucleoti
c	30	14	56.0	5773	21	AAC76014	Human ORFX ORF1569
c	31	14	56.0	5876	22	AAI59303	Human polynucleoti
	32	14	56.0	32351	21	AAF21307	Human low adenosin
	33	14	56.0	32351	21	AAA35185	Human adenosine re
	34	14	56.0	40298	21	AAF21311	Human low adenosin
	35	14	56.0	40298	21	AAA35189	Human adenosine re
	36	14	56.0	1038602	20	AAZ01425	Complete genome se
	37	13	52.0	87	16	AAT22576	Human gene signatu
c	38	13	52.0	167	16	AAT20233	Human gene signatu
c	39	13	52.0	187	21	AAC26213	Human secreted pro
c	40	13	52.0	363	21	AAA50159	Antibody 5H7 heavy
c	41	13	52.0	378	13	AAQ25667	Sequence of the an
c	42	13	52.0	408	21	AAA74604	Murine 5B3 antibod
c	43	13	52.0	415	21	AAC56783	Eucalyptus grandis
c	44	13	52.0	476	21	AAC09392	Human secreted pro
	45	13	52.0	556	22	AAH09733	Human cDNA clone (
c	46	13	52.0	559	22	AAF93531	cDNA encoding SRT
	47	13	52.0	603	22	AAH32562	Human secreted pro
	48	13	52.0	632	22	AAH97876	Murine 7-transmemb
	49	13	52.0	637	21	AAA95830	Bovine metalloprot
c	50	13	52.0	713	15	AAQ81957	CZP2(487-713). Ca
c	51	13	52.0	719	22	AAD04227	Bovine rotavirus s
c	52	13	52.0	735	22	AAH19684	Murine anti-human
c	53	13	52.0	777	18	AAT79355	Staphylococcus aur
c	54	13	52.0	778	20	AAX32456	Partial nucleotide
c	55	13	52.0	902	21	AAZ56756	Human transmembran
c	56	13	52.0	930	20	AAX15335	cDNA encoding a le
c	57	13	52.0	930	22	AAH32490	Human olfactory re
c	58	13	52.0	1002	21	AAA50164	DNA encoding 5H7 s
	59	13	52.0	1010	21	AAC64778	Strawberry alcohol
	60	13	52.0	1010	21	AAA27678	Strawberry alcohol
c	61	13	52.0	1278	21	AAC36039	Arabidopsis thalia
c	62	13	52.0	1317	21	AAA94840	Drosophila melanog
c	63	13	52.0	1359	21	AAA72281	Drosophila odorant
c	64	13	52.0	1404	20	AAZ31521	Consensus phytase-
c	65	13	52.0	1404	21	AAA73234	Consensus phytase
c	66	13	52.0	1404	21	AAA73292	Consensus phytase
c	67	13	52.0	1404	21	AAA73293	Consensus phytase
c	68	13	52.0	1404	21	AAZ59716	DNA encoding a mut
c	69	13	52.0	1404	22	AAS05823	DNA encoding conse
c	70	13	52.0	1410	21	AAC59955	Human secreted pro
c	71	13	52.0	1426	21	AAA73232	Consensus phytase
c	72	13	52.0	1426	21	AAZ59642	DNA encoding phyta
c	73	13	52.0	1426	22	AAS05819	DNA encoding conse
c	74	13	52.0	1490	21	AAA16645	Human secreted pro
	75	13	52.0	1594	22	AAH32526	Human secreted pro
c	76	13	52.0	1600	22	AAI59481	Human polynucleoti
c	77	13	52.0	1701	22	AAS21258	Human cDNA sequenc
c	78	13	52.0	1713	21	AAA40495	Human fetal kidney
c	79	13	52.0	1726	20	AAX07570	Homo sapiens fetal
c	80	13	52.0	1887	19	AAT94207	AEPII1a clone 63GP
c	81	13	52.0	1903	22	AAD05205	Drosophila melanog
c	82	13	52.0	1929	21	AAZ93789	Human MDG1 gene (C
c	83	13	52.0	2044	21	AAZ43754	Murine NNX3 cDNA.
	84	13	52.0	2058	22	AAH14074	Human cDNA sequenc

c 85	13	52.0	2214	15	AAQ65604	Porcine zona pellu
c 86	13	52.0	2214	20	AAZ22709	Porcine zona pellu
c 87	13	52.0	2214	20	AAV64786	Porcine ZPA cDNA.
c 88	13	52.0	2214	21	AAZ95646	Porcine zona pellu
c 89	13	52.0	2214	21	AAZ46254	Porcine oocyte zon
c 90	13	52.0	2214	21	AAZ33243	Porcine zona pellu
c 91	13	52.0	2214	21	AAZ37798	Porcine zona pellu
c 92	13	52.0	2216	15	AAQ70072	Canine zona pelluc
93	13	52.0	2277	22	AAF61089	P. putida KT2440-a
c 94	13	52.0	2381	15	AAQ65608	Canine zona pelluc
c 95	13	52.0	2381	20	AAZ22713	Canine zona pelluc
c 96	13	52.0	2381	20	AAV64790	Canine ZPA cDNA.
c 97	13	52.0	2381	21	AAZ95650	Canine zona pelluc
c 98	13	52.0	2381	21	AAZ46258	Canine oocyte zona
c 99	13	52.0	2381	21	AAZ33247	Canine zona pelluc
c 100	13	52.0	2381	21	AAZ37802	Canine zona pelluc

ALIGNMENTS

RESULT 1
AAI62941
ID AAI62941 standard; DNA; 839 BP.
XX
AC AAI62941;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human genomic DNA SEQ ID NO 269.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200155449-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01346.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0218290.
PR 14-AUG-2000; 2000US-0225447.
PR 01-SEP-2000; 2000US-0229343.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231243.
PR 25-SEP-2000; 2000US-0234997.
PR 29-SEP-2000; 2000US-0236367.
PR 13-OCT-2000; 2000US-0239937.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246528.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249265.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476225/51.
XX
PT Novel plasma membrane associated proteins useful for diagnosing,
PT treating, preventing and/or prognosing disorders related to the
PT proteins, including cancer, immune response and neuronal disorders -
XX
PS Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;
Best Local Similarity 100.0%; Pred. No. 9.9e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 gacattgtcgcaagtcacagaatta 25
          |||||||||||||||||||||||||
Db     53 gacattgtcgcaagtcacagaatta 77
```

RESULT 2
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX

```
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Patton DA;
XX
DR   WPI; 1999-152902/13.
DR   P-PSDB; AAW73906.
XX
PT   Transgenic plants with high biotin levels - transformed with DNA
PT   encoding di:amino-pelargonic acid amino-transferase or biotin
PT   synthase
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db      21 gacattgtcgcaagtcacagaatta 45


RESULT   3
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db      44 gacattgtcgcaagtcacagaatta 68


RESULT   4
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzym.
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db      62 gacattgtcgcaagtcacagaatta 86


RESULT   5
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
```

```
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag=  a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag=  b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag=  c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag=  d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag=  e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag=  f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag=  g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag=  h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag=  i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag=  j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag=  k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag=  l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag=  m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag=  n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag=  o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag=  p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag=  q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag=  r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 9.9e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db     137 gacattgtcgcaagtcacagaatta 161


RESULT   6
AAQ56307/c
ID   AAQ56307 standard; DNA; 17 BP.
XX
AC   AAQ56307;
```

```
XX
DT   13-APR-1994  (first entry)
XX
DE   BioB DNA primer.
XX
KW   BioA; BioB; promoter; biotin; operon; primer; ss.
XX
OS   Synthetic.
XX
PN   JP05219956-A.
XX
PD   31-AUG-1993.
XX
PF   14-SEP-1992;   92JP-0244792.
XX
PR   14-SEP-1992;   92JP-0244792.
XX
PA   (SHIS ) SHISEIDO CO LTD.
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
DR   WPI; 1993-308323/39.
XX
PT   DNA sequence of biotin operon - has base sequence of E. coli
PT   mutated by base pair(s) compared to wild type
XX
PS   Example 1; Fig 8; 11pp; Japanese.
XX
CC   A novel DNA sequence comprises the E.coli biotin operon (BO) in which
CC   the control region of BO or the region near the bioB initiation
CC   codon is mutated by at least one base pair compared to its
CC   wild type.  Two primers (AAQ56306-Q56307) are described in Example 1.
CC   A microorganism belonging to Escherichia genus, transformed by
CC   a recombinant plasmid carrying such DNA can be used for the prodn.
CC   of biotin-active substances.
XX
SQ   Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;
```

```
  Query Match             68.0%;  Score 17;  DB 14;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 0.47;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gtcgcaagtcacagaat 23
           |||||||||||||||||
Db      17 GTCGCAAGTCACAGAAT 1
```

```
RESULT   7
AAF94235
ID   AAF94235 standard; DNA; 561 BP.
XX
AC   AAF94235;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Primer specific for DNA encoding secretory/membrane protein SEQ ID 669.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; PCR primer; ss.
XX
OS   Synthetic.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 5; SEQ ID 669; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 561 BP; 213 A; 83 C; 114 G; 149 T; 2 other;
```

```
  Query Match             60.0%;  Score 15;  DB 22;  Length 561;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     339 caagtcacagaatta 353
```

```
RESULT   8
AAZ30602
ID   AAZ30602 standard; DNA; 1771 BP.
XX
AC   AAZ30602;
XX
DT   18-JAN-2000  (first entry)
XX
DE   Mouse integrin alpha 4 coding sequence.
XX
KW   Human; integrin; antisense; oligonucleotide; inhibition; expression;
KW   very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;
KW   vascular endothelial cell; vascular endothelium; migration; inflammation;
KW   atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
KW   metastasis; circulatory system; autoimmune disease; Grave's disease;
KW   Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.
XX
OS   Mus sp.
XX
PN   US5968826-A.
XX
PD   19-OCT-1999.
XX
PF   05-OCT-1998;   98US-0166203.
XX
PR   05-OCT-1998;   98US-0166203.
```

```
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Bennett CF,  Cowsert LM,  Condon TP;
XX
DR   WPI; 1999-590416/50.
DR   P-PSDB; AAY44154.
XX
PT   Antisense inhibition of integrin alpha4 expression useful for treating
PT   inflammatory diseases such as atherosclerosis, allergies, asthma and
PT   arthritis -
XX
PS   Example 12; Column 71-74; 40pp; English.
XX
CC   This sequence represents the coding region of the mouse integrin
CC   alpha4 gene. The invention relates to the generation of antisense
CC   oligonucleotides targeted to the integrin alpha4 gene which are used
CC   for inhibiting expression of the integrin alpha4 mRNA or protein.
CC   Integrin alpha4 is a component of Very Late Antigen (VLA)-4 (also
CC   called alpha4beta1 and CD49d/CD29). VLA-4 is expressed on the cell
CC   surfaces of leucocytes and vascular endothelial cells and mediates the
CC   adhesion of leucocytes to the vascular endothelium prior to migration
CC   into the surrounding tissues. This migration is an essential step in
CC   inflammation and hence VLA-4 (and consequently integrin alpha4) is a
CC   potential therapeutic target for treating inflammatory diseases and
CC   the damaging effects of excessive inflammation. These disorders include
CC   atherosclerosis, allergies, asthma, rheumatoid arthritis and tumor cell
CC   metastasis (VLA-4 is involved in migration of the tumor cells through
CC   the extracellular matrix into the circulatory system). VLA-4 is also
CC   involved in a number of autoimmune diseases such as Grave's disease,
CC   Hashimoto's thyroiditis, encephalomyelitis (EAE), multiple sclerosis.
CC   VLA-4 may also be involved in promoting adhesion (i.e. retention) of
CC   hemopoietic stem cells in bone-marrow and in allograft rejection.
XX
SQ   Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 20;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT   9
AAS10264
ID   AAS10264 standard; DNA; 1771 BP.
XX
AC   AAS10264;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Mouse integrin alpha 4 partial gene sequence.
XX
KW   Mouse; integrin alpha 4; antisense; L20788; very late antigen 4; VLA4;
KW   autoimmune disease; inflammatory disease; rheumatoid arthritis;
KW   multiple sclerosis; tumour metastasis; melanoma; asthma; psoriasis;
KW   allergy; Grave's disease; Hashimoto's thyroiditis;
KW   systemic lupus erythematosus; allograft rejection; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1193..1771
FT                   /*tag=  a
FT                   /product= "Integrin alpha 4"
FT                   /partial
FT                   /note= "No stop codon"
FT   intron          1387..1708
FT                   /*tag=  b
FT                   /number= 1
XX
PN   US6258790-B1.
XX
PD   10-JUL-2001.
XX
PF   19-AUG-1999;   99US-0377309.
XX
PR   05-OCT-1998;   98US-0166203.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Bennett CF,  Condon TP,  Cowsert LM;
XX
DR   WPI; 2001-450381/48.
DR   P-PSDB; AAU04967.
XX
PT   Composition for treating inflammatory and autoimmune diseases,
PT   comprises antisense compound targeted to nucleic acid molecule encoding
PT   integrin alpha4 and inhibit expression of integrin alpha4 -
XX
PS   Example 12; Column 77-79; 49pp; English.
XX
CC   The sequence (Genbank accession number L20788) is the partial gene
CC   sequence of mouse integrin alpha 4, one component of very late antigen 4
CC   (VLA4) which is involved in various autoimmune and inflammatory diseases.
CC   The invention relates to antisense inhibitors of integrin alpha 4 which
CC   target and inhibit expression of integrin alpha 4. The antisense
CC   molecules are useful for inhibiting the expression of integrin alpha4 in
CC   human cells or tissues, treating an animal having a disease or
CC   condition associated with expression of integrin alpha4, e.g.,
CC   inflammatory disease or condition, autoimmune disease or condition
CC   including rheumatoid arthritis, multiple sclerosis and tumour
CC   metastases, melanoma, asthma, psoriasis, allergy, Grave's disease,
CC   Hashimoto's thyroiditis, systemic lupus erythematosus and allograft
CC   rejection, and diseases or conditions characterised by leukocyte
CC   migration into affected tissues, preferably central nervous system
CC   tissues. The antisense molecules are also useful for reducing the
CC   levels of VLA-4 and alpha4beta7 integrin in human cells or tissues,
CC   and reducing the adherence of cells of a first type e.g., melanoma cells
CC   or lymphocytes, to cells of a second type e.g., endothelial cells, by
CC   inhibiting integrin alpha4 expression and thus decreasing adhesion of
CC   cells.
XX
SQ   Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 22;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT  10
AAF93909/c
ID   AAF93909 standard; cDNA; 3520 BP.
XX
AC   AAF93909;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Human cDNA encoding a membrane or secretory protein clone PSEC0256.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
```

```
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
DR   P-PSDB; AAB88482.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 1; SEQ ID 331; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;

  Query Match             60.0%;  Score 15;  DB 22;  Length 3520;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    3182 CAAGTCACAGAATTA 3168


RESULT  11
AAF29361/c
ID   AAF29361 standard; DNA; 3520 BP.
XX
AC   AAF29361;
XX
DT   20-APR-2001  (first entry)
XX
DE   Amyloid-beta protein agglutination regulating factor DNA SEQ ID 9.
XX
KW   Human; amyloid-beta protein; agglutination regulatory factor;
KW   Alzheimer's disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200104299-A1.
XX
PD   18-JAN-2001.
XX
PF   06-JUL-2000; 2000WO-JP04515.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   18-OCT-1999;   99US-0159586.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Yamazaki M,  Satoh S;
PI   Arakawa H,  Morita M;
XX
DR   WPI; 2001-138347/14.
DR   P-PSDB; AAB49771.
XX
PT   Polynucleotide encoding Amyloid-beta protein agglutination-controlling
PT   factor, useful for inhibiting or promoting agglutination or
PT   sedimentation of amyloid-beta protein and in diagnosis and screening
PT   drugs for Alzheimer's disease -
XX
PS   Claim 1; Page 58-63; 72pp; Japanese.
XX
CC   This invention relates to polynucleotides AAF29357 - AAF29361 which
CC   encode proteins AAB49767 - AAB49771. The proteins inhibit or promote the
CC   agglutination of amyloid beta protein. The protein and polynucleotide
CC   sequences are useful in the diagnosis of Alzheimer's disease. They are
CC   also useful for screening drugs which are useful for treating Alzheimer's
CC   disease.
XX
SQ   Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;

  Query Match             60.0%;  Score 15;  DB 22;  Length 3520;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    3182 CAAGTCACAGAATTA 3168


RESULT  12
AAF98729/c
ID   AAF98729 standard; DNA; 3748 BP.
XX
AC   AAF98729;
XX
DT   02-JUL-2001  (first entry)
XX
DE   Human late stage ovarian tumour polynucleotide marker 44.
XX
KW   Human; ovarian cancer; identification; detection; characterisation;
KW   tumour; kinase; marker; cytostatic; antisense gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200118542-A2.
XX
PD   15-MAR-2001.
XX
PF   01-SEP-2000; 2000WO-US24199.
XX
PR   03-SEP-1999;   99US-0152547.
PR   16-MAR-2000; 2000US-0190347.
PR   21-MAR-2000; 2000US-0191321.
PR   31-MAY-2000; 2000US-0208382.
PR   20-JUL-2000; 2000US-0220467.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lee J,  Thompsho P,  Lillie J;
XX
DR   WPI; 2001-211428/21.
```

```
XX
PT   Detection, assessment, prevention and therapy of ovarian cancer,
PT   comprises detecting changes in the expression of a variety of markers -
XX
PS   Claim 1; Page 1196-1197; 1198pp; English.
XX
CC   The present invention describes a method for assessing whether a patient
CC   is afflicted with ovarian cancer by comparing: (1) the expression of a
CC   marker (I) (see AAF98594 to AAF98730), in a patient sample; and (2) the
CC   normal level of expression of (I) in a control non-ovarian cancer
CC   sample, where a significant difference between the level of expression
CC   in (a) and (b) is an indication that the patient is afflicted with
CC   ovarian cancer. (I) have cytostatic activities and can be used in
CC   antisense gene therapy. The method, compositions and kits from the
CC   present invention can be used for: (1) assessing and treating ovarian
CC   cancer; (2) making isolated hybridoma, which produces an antibody useful
CC   for ovarian cancer assessment; and (3) inhibiting ovarian cancer in a
CC   patient. AAF98573 to AAF98593 represent human kinase marker primers and
CC   probes which are used in the exemplification of the present invention.
XX
SQ   Sequence 3748 BP; 1110 A; 825 C; 761 G; 1052 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 22;  Length 3748;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db    2627 CAAGTCACAGAATTA 2613


RESULT  13
AAS22683
ID   AAS22683 standard; cDNA; 6087 BP.
XX
AC   AAS22683;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human cDNA encoding a novel human protein #249.
XX
KW   Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW   immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW   anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW   antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW   thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW   Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW   tissue regeneration; immune disorder.
XX
OS   Homo sapiens.
XX
PN   WO200155437-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02623.
XX
PR   25-JAN-2000; 2000US-0491404.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-451939/48.
DR   P-PSDB; AAU14378.
XX
PT   Isolated polypeptides useful for treating anti-inflammatory diseases,
PT   nervous system disorders, and for regenerating bone and cartilage -
XX
PS   Claim 1; Page 676-677; 894pp; English.
XX
CC   The invention relates to polynucleotides encoding novel human
CC   proteins or their active domains. The polypeptides, polynucleotides and
CC   antibodies raised against the polypeptides are used in a method of
CC   treatment of a mammal and prevention of disorders caused by the aberrant
CC   protein expression or activity.  The polypeptides can be used as
CC   molecular weight markers, food supplements, and in antibody production.
CC   The polypeptides are used to identify compounds which bind to the
CC   polypeptides. Polynucleotides of the invention are used as probes and
CC   primers, for sequencing, for chromosome or gene mapping, in the
CC   production of recombinant proteins, and in generating anti-sense DNA or
CC   RNA and in gene therapy.  Polypeptides of the invention can be used to
CC   target drugs to a tumour, in assays to determine biological activity, to
CC   raise antibodies/elicit an immune response, to determine quantitative
CC   protein levels, as tissue markers, and to isolate receptors or ligands.
CC   Polypeptides of the invention may also be useful in treating platelet
CC   disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC   ligament and/or nerve tissue, wound healing, treating burns, promoting
CC   the proliferation, differentiation and survival of stem cells, as a
CC   contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC   Alzheimer's, Parkinson's and Huntington's diseases, amylotrophic lateral
CC   sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC   fungal infection or from autoimmunity, cancer, allergy, asthma,
CC   graft-versus-host disease, eczema, haemophilia, thrombosis,
CC   anti-inflammatory diseases, nervous system disorders, and infection.
CC   The present sequence encodes a protein of the invention.
XX
SQ   Sequence 6087 BP; 1490 A; 1609 C; 1574 G; 1413 T; 1 other;


  Query Match             60.0%;  Score 15;  DB 22;  Length 6087;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 tcgcaagtcacagaa 22
           ||||||||||||||||
Db    5191 tcgcaagtcacagaa 5205


RESULT  14
AAS22447
ID   AAS22447 standard; cDNA; 6116 BP.
XX
AC   AAS22447;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human cDNA encoding a novel human protein #13.
XX
KW   Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW   immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW   anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW   antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW   thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW   Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW   tissue regeneration; immune disorder.
XX
OS   Homo sapiens.
XX
PN   WO200155437-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US02623.
XX
PR   25-JAN-2000; 2000US-0491404.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-451939/48.
DR   P-PSDB; AAU14142.
XX
PT   Isolated polypeptides useful for treating anti-inflammatory diseases,
```

```
PT   nervous system disorders, and for regenerating bone and cartilage -
XX
PS   Claim 1; Page 190-196; 894pp; English.
XX
CC   The invention relates to polynucleotides encoding novel human
CC   proteins or their active domains. The polypeptides, polynucleotides and
CC   antibodies raised against the polypeptides are used in a method of
CC   treatment of a mammal and prevention of disorders caused by the aberrant
CC   protein expression or activity. The polypeptides can be used as
CC   molecular weight markers, food supplements, and in antibody production.
CC   The polypeptides are used to identify compounds which bind to the
CC   polypeptides. Polynucleotides of the invention are used as probes and
CC   primers, for sequencing, for chromosome or gene mapping, in the
CC   production of recombinant proteins, and in generating anti-sense DNA or
CC   RNA and in gene therapy. Polypeptides of the invention can be used to
CC   target drugs to a tumour, in assays to determine biological activity, to
CC   raise antibodies/elicit an immune response, to determine quantitative
CC   protein levels, as tissue markers, and to isolate receptors or ligands.
CC   Polypeptides of the invention may also be useful in treating platelet
CC   disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC   ligament and/or nerve tissue, wound healing, treating burns, promoting
CC   the proliferation, differentiation and survival of stem cells, as a
CC   contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC   Alzheimer's, Parkinson's and Huntington's diseases, amylotrophic lateral
CC   sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC   fungal infection or from autoimmunity, cancer, allergy, asthma,
CC   graft-versus-host disease, eczema, haemophilia, thrombosis,
CC   anti-inflammatory diseases, nervous system disorders, and infection.
CC   The present sequence encodes a protein of the invention.
XX
SQ   Sequence 6116 BP; 1491 A; 1626 C; 1582 G; 1417 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 22;  Length 6116;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 tcgcaagtcacagaa 22
           |||||||||||||||
Db    5220 tcgcaagtcacagaa 5234


RESULT  15
AAQ03324/c
ID   AAQ03324 standard; DNA; 6567 BP.
XX
AC   AAQ03324;
XX
DT   02-AUG-1990  (first entry)
XX
DE   Eimeria tenella genomic DNA encoding antigen GX5401FL.
XX
KW   Eimeria tenella; antigen GX5401FL; antigen GX5401; avian coccidiosis.
XX
OS   Eimeria tenella.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..6567
FT                   /*tag=  a
XX
PN   WO9000403-A.
XX
PD   25-JAN-1990.
XX
PF   05-JUL-1989;   89WO-US02918.
XX
PR   05-JUL-1988;   88US-0215162.
XX
PA   (GENE-) GENEX CORP.
XX
PI   Anderson DM,  McCandliss RJ,  Strausberg SL,  Strausberg RL;
XX
DR   WPI; 1990-051586/07.
DR   P-PSDB; AAR05222.
XX
PT   Cloned gene or fragment encoding antigenic protein -
PT   which binds with antibodies against avian coccidia, and
PT   transformed cells used in vaccine
XX
PS   Claim 10; Page 93 and Fig 14; 134pp; English.
XX
CC   An E. tenella genomic library was screened with radioactively labeled
CC   cDNA encoding the GX5401 antigen to identify clones encoding extensions
CC   of the GX5401 coding sequence. Several plaques reactive with the cDNA
CC   were identified. DNA from clone number 533 was sequenced. This analysis
CC   showed that the coding sequence in this clone overlaps with the GX5401
CC   coding sequence and extends that sequence toward the N-terminal coding
CC   region. Analysis of the new sequence together with the GX5401 antigen
CC   coding sequence reveals an open reading frame encoding an Eimeria
CC   protein of about 250,000 daltons. The protein carries several repeated
CC   peptide sequences and is rich in cysteine residues. The open reading
CC   frame encodes a potential signal sequence for protein secretion. The
CC   coding sequence for this antigenic protein is given in AAN93324. Also new
CC   are an expression vector contg. the cloned gene, host cells transformed
CC   with the vector, and AP encoded by the cloned gene. The transformed cells
CC   are used in a vaccine to immunise birds against avian coccidiosis. By
CC   labelling the peptides, they can be used as a type-specific probe. The
CC   AP may also be used in an assay to detect Ab against the coccidia. The
CC   Abs are used to identify transformed cells contg. the DNA.
XX
SQ   Sequence 6567 BP; 1490 A; 1741 C; 2189 G; 1147 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 11;  Length 6567;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagt 15
           |||||||||||||||
Db    3486 GACATTGTCGCAAGT 3472


RESULT  16
AAH12853
ID   AAH12853 standard; cDNA; 523 BP.
XX
AC   AAH12853;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:9688.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000;  2000JP-0118776.
PR   02-MAY-2000;  2000JP-0183767.
PR   09-JUN-2000;  2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
```

PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3; SEQ ID 9688; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 523 BP; 206 A; 87 C; 70 G; 151 T; 9 other;

Query Match 56.0%; Score 14; DB 22; Length 523;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatt 24
||||||||||||||
Db 390 caagtcacagaatt 403

RESULT 17
AAN80927/c
ID AAN80927 standard; DNA; 1050 BP.
XX
AC AAN80927;
XX
DT 15-OCT-1990 (first entry)
XX
DE Lividans Exported Protein gene.
XX
KW Longisporus trypsin inhibitor; Lividans exported protein; LTI; LEP-10;
KW protease inhibitor; probe; ss.
XX
OS Streptomyces lividans strain 1326.
XX
FH Key Location/Qualifiers
FT sig_peptide 394..498
FT /*tag= a
FT mat_peptide 499.. 825
FT /*tag= b
FT /product=LEP-10
XX
PN WO8801278-A.
XX
PD 25-FEB-1988.
XX
PF 17-AUG-1987; 87WO-US02009.
XX
PR 18-AUG-1986; 86US-0897245.
XX
PA (SMIK) SMITHKLINE BECKMAN CORP.
PA (BERK/) TR BERKA.
XX
P1 Berka TR, Fornwald JA, Gorniak JG, Rosenberg M;
XX
DR WPI; 1988-063996/09.
DR P-PSDB; AAP80944.
XX
PT Novel protease inhibitors LEP-10 and LTI - produced by Streptomyces spp.
XX
PS Disclosure; ; pp; English.
XX
CC The sequence can be cloned and used to produce recombinant LEP-10.
CC The DNA can be used as a probe for identifying further coding
CC sequences for protease inhibitors; the protein is useful as an
CC assay reagent.
CC See also AAN80926.
XX
SQ Sequence 1050 BP; 149 A; 398 C; 366 G; 137 T; 0 other;

Query Match 56.0%; Score 14; DB 9; Length 1050;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 aagtcacagaatta 25
||||||||||||||
Db 251 AAGTCACAGAATTA 238

RESULT 18
AAH14046/c
ID AAH14046 standard; cDNA; 1465 BP.
XX
AC AAH14046;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:11168.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11168; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the

```
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1465 BP; 470 A; 259 C; 270 G; 466 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 1465;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatt 24
           ||||||||||||||
Db    1076 CAAGTCACAGAATT 1063


RESULT  19
AAT31932/c
ID   AAT31932 standard; DNA; 1677 BP.
XX
AC   AAT31932;
XX
DT   27-SEP-1996  (first entry)
XX
DE   Retinoid X receptor interacting protein RIP14-2 DNA.
XX
KW   Retinoid X receptor interacting protein; RXR; RIP; RIP14-2; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1356
FT                   /*tag=  a
FT   polyA_signal    1672-1677
FT                   /*tag=  b
XX
PN   WO9621677-A1.
XX
PD   18-JUL-1996.
XX
PF   08-DEC-1995;   95WO-US16311.
XX
PR   13-JAN-1995;   95US-0372652.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Choi H,  Moore D,  Seol W;
XX
DR   WPI; 1996-342241/34.
DR   P-PSDB; AAR99739.
XX
PT   Retinoid X receptor (RXR) interacting protein (RIP)  - useful to
PT   modulate or mediate RXR function, anti-RIP antibodies can be used to
PT   determine RIP subcellular distribution patterns
XX
PS   Claim 8; Page 59-60; 90pp; English.
XX
CC   A cDNA clone (AAT31932) codes for retinoid X receptor interacting
CC   protein RIP14-2 (AAR99739), a previously undescribed orphan member
CC   of the nuclear receptor superfamily which may be involved in the
CC   complex retinoid response.  It was isolated from a mouse liver
CC   cDNA library using an RIP14 probe.  Different RIP14-2 clones had
CC   different 5' sequences (see also AAT31935-37).  Another isoform,
CC   RIP14-1, was also detected (see also AAT31928).  The cDNA can be
CC   used for the recombinant prodn. of RIP14-2 in transformed host
CC   cells.
XX
SQ   Sequence 1677 BP; 486 A; 385 C; 408 G; 398 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 17;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     630 TTGTCGCAAGTCAC 617


RESULT  20
AAT31928/c
ID   AAT31928 standard; DNA; 1787 BP.
XX
AC   AAT31928;
XX
DT   27-SEP-1996  (first entry)
XX
DE   Retinoid X receptor interacting protein RIP14-1 DNA.
XX
KW   Retinoid X receptor interacting protein; RXR; RIP; RIP14-1; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             12..1466
FT                   /*tag=  a
FT   polyA_signal    1782..1787
FT                   /*tag=  b
XX
PN   WO9621677-A1.
XX
PD   18-JUL-1996.
XX
PF   08-DEC-1995;   95WO-US16311.
XX
PR   13-JAN-1995;   95US-0372652.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Choi H,  Moore D,  Seol W;
XX
DR   WPI; 1996-342241/34.
DR   P-PSDB; AAR99735.
XX
PT   Retinoid X receptor (RXR) interacting protein (RIP)  - useful to
PT   modulate or mediate RXR function, anti-RIP antibodies can be used to
PT   determine RIP subcellular distribution patterns
XX
PS   Claim 8; Page 53-54; 90pp; English.
XX
CC   A full-length cDNA clone (AAT31928) codes for retinoid X receptor
CC   interacting protein RIP14-1 (AAR99735), a previously undescribed
CC   orphan member of the nuclear receptor superfamily which may be
CC   involved in the complex retinoid response.  It was isolated from
CC   a mouse liver cDNA library using an RIP14 probe.  Another isoform,
CC   RIP14-2, was also detected (see also AAT31932).  The cDNA can be used
CC   for the recombinant prodn. of RIP14-1 in transformed host cells.
XX
SQ   Sequence 1787 BP; 515 A; 408 C; 442 G; 422 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 17;  Length 1787;
```

```
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     740 TTGTCGCAAGTCAC 727


RESULT  21
AAZ52940/c
ID   AAZ52940 standard; cDNA; 1890 BP.
XX
AC   AAZ52940;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Human prostate tumor cDNA library derived EST fragment #83.
XX
KW   Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW   treatment; ds.
XX
OS   Homo sapiens.
XX
PN   DE19820190-A1.
XX
PD   04-NOV-1999.
XX
PF   28-APR-1998;   98DE-1020190.
XX
PR   28-APR-1998;   98DE-1020190.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-621386/54.
DR   P-PSDB; AAY74061, AAY74062, AAY74063.
XX
PT   New human nucleic acid sequences from pancreatic tumors, and related
PT   proteins  -
XX
PS   Claim 2; Page 250; 502pp; German.
XX
CC   This invention describes novel polypeptides and their encoding nucleic
CC   acids derived from human pancreatic tumor tissue which have cytostatic
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. AAZ52858-Z53014
CC   represent expressed sequence tag (EST) fragments derived from a human
CC   pancreatic tumor cDNA library and which encode the proteins represented
CC   in AAY73814-Y74252.
XX
SQ   Sequence 1890 BP; 590 A; 358 C; 398 G; 544 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 1890;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatt 24
           ||||||||||||||
Db    1475 CAAGTCACAGAATT 1462


RESULT  22
AAI60385
ID   AAI60385 standard; cDNA; 1919 BP.
XX
AC   AAI60385;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4374.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM41229.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 4374; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1919 BP; 564 A; 394 C; 343 G; 617 T; 1 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 1919;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatt 24
           ||||||||||||||
Db    402 caagtcacagaatt 415


RESULT  23
AAH15029/c
ID   AAH15029 standard; cDNA; 2022 BP.
XX
AC   AAH15029;
XX
```

```
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:13001.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 13001; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2022 BP; 622 A; 389 C; 418 G; 593 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 2022;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatt 24
           ||||||||||||||
Db    1633 CAAGTCACAGAATT 1620


RESULT  24
AAI58599/c
ID   AAI58599 standard; cDNA; 2102 BP.
XX
AC   AAI58599;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 802.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM39443.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 802; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 2102 BP; 662 A; 390 C; 441 G; 609 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 2102;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatt 24
           ||||||||||||||
Db    1564 CAAGTCACAGAATT 1551
```

```
RESULT  25
AAD12786/c
ID   AAD12786 standard; cDNA; 2220 BP.
XX
AC   AAD12786;
XX
DT   23-OCT-2001  (first entry)
XX
DE   Human nuclear hormone receptor (NHREC)-3 cDNA.
XX
KW   Human; nuclear hormone receptor; NHREC-3; intracellular receptor; cancer;
KW   transcriptional regulator; metabolic disorder; lysosomal storage disease;
KW   Addison's disease; cystic fibrosis; endocrine disorder; atherosclerosis;
KW   breast cancer; osteoporosis; Cushing's disease; reproductive disorder;
KW   developmental disorder; infertility; renal tubular acidosis; cirrhosis;
KW   endometriosis; hereditary neuropathy; immunological disorder; anaemia;
KW   acquired immunodeficiency syndrome; AIDS; cell proliferative disorder;
KW   allergy; Grave's disease; gene therapy; drug screening; cytostatic;
KW   immunomodulatory; gout; diabetes; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             337..1770
FT                   /*tag= a
FT                   /product= "Human nuclear hormone receptor (NHREC)-3"
XX
PN   WO200155392-A2.
XX
PD   02-AUG-2001.
XX
PF   26-JAN-2001; 2001WO-US02741.
XX
PR   28-JAN-2000; 2000US-0178578.
PR   25-FEB-2000; 2000US-0185079.
PR   17-MAR-2000; 2000US-0189999.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Burford N, Baughn MR, Tang YT, Bandman O, Tribouley CM, Das D;
PI   Policky JL, Lo TP;
XX
DR   WPI; 2001-488797/53.
DR   P-PSDB; AAE06680.
XX
PT   New human nuclear hormone receptors and polynucleotides encoding the
PT   receptors, useful for diagnosing, treating and preventing metabolic,
PT   endocrine, reproductive, developmental, immunological and cell
PT   proliferative disorders  -
XX
PS   Claim 5; Page 95-96; 99pp; English.
XX
CC   The present sequence is human nuclear hormone receptor (NHREC)-3 cDNA.
CC   NHREC known as nuclear receptor or intracellular receptor, constitute
CC   a protein superfamily whose members are both receptors and
CC   transcriptional regulators. NHREC is useful in the diagnosis, treatment
CC   and prevention of metabolic (e.g. Addison's disease, cystic fibrosis,
CC   diabetes, lysosomal storage disease), endocrine (e.g. disorders of
CC   pituitary, thyroid, Cushing's disease, osteoporosis), reproductive (e.g.
CC   infertility, endometriosis, breast cancer), developmental (e.g. anaemia,
CC   renal tubular acidosis, hereditary neuropathy), immunological disorders
CC   (e.g. acquired immunodeficiency syndrome (AIDS), allergy, Grave's
CC   disease, gout) and cell proliferative disorders (e.g. atherosclerosis,
CC   cirrhosis, cancer) and in the assessment of the effects of exogenous
CC   compounds on the expression of nucleic acid and amino acid sequences of
CC   nuclear hormone receptors. NHREC may also be used to screen for
CC   compounds that modulate the activity of NHREC, for somatic or germline
CC   gene therapy, to detect single nucleotide polymorphisms, as element on
CC   microarray, to generate hybridisation probes and in drug screening.
XX
SQ   Sequence 2220 BP; 710 A; 425 C; 465 G; 620 T; 0 other;

  Query Match             56.0%;  Score 14;  DB 22;  Length 2220;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db    1041 TTGTCGCAAGTCAC 1028


RESULT  26
AAV10120/c
ID   AAV10120 standard; cDNA; 2233 BP.
XX
AC   AAV10120;
XX
DT   29-MAY-1998  (first entry)
XX
DE   Human retinoid receptor RR1 cDNA.
XX
KW   Retinoid receptor; RR1; steroid receptor; agonist; antagonist; cancer;
KW   adrenal deficiency; skin disorder; inflammatory disorder;
KW   immune response regulator; autoimmune disease; therapeutic antibody; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             363..1781
FT                   /*tag= a
FT                   /product= RR1
FT                   /note= "retinoid receptor"
XX
PN   US5728548-A.
XX
PD   17-MAR-1998.
XX
PF   29-JUN-1995;   95US-0496631.
XX
PR   29-JUN-1995;   95US-0496631.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Bowman M;
XX
DR   WPI; 1998-206567/18.
DR   P-PSDB; AAW40072.
XX
PT   Human retinoid receptor protein RR1 - useful for, e.g. drug
PT   screening, therapy and antibody production
XX
PS   Claim 8; Column 11-15; 13pp; English.
XX
CC   This sequence encodes a novel human steroid receptor, the retinoid
CC   receptor protein or RR1. This protein can be used in screening assays
CC   for steroid hormone receptor agonists and antagonists and in
CC   pharmaceutical compositions for treating adrenal deficiencies, e.g.
CC   Addison's disease, cancer, skin disorders, e.g. acne and psoriasis,
CC   inflammatory disorders, e.g. arthritis and HIV infections. The protein
CC   can also be used for regulating immune responses, e.g. as antitumour
CC   agents, vaccine adjuvants, organ rejection inhibitors or agents for
CC   treating autoimmune diseases. The protein can further be used to produce
CC   therapeutic antibodies.
XX
SQ   Sequence 2233 BP; 744 A; 427 C; 463 G; 599 T; 0 other;

  Query Match             56.0%;  Score 14;  DB 19;  Length 2233;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db    1055 TTGTCGCAAGTCAC 1042
```

```
RESULT  27
AAF77541/c
ID   AAF77541 standard; DNA; 2585 BP.
XX
AC   AAF77541;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Caenorhabditis briggsae mab-21 coding sequence.
XX
KW   Mab-21; cell fate choice; pax-6; aniridia; Moebius syndrome;
KW   chromosome 13q13; peripheral neural tissue differentiation; ds.
XX
OS   Caenorhabditis briggsae.
XX
PN   US6197504-B1.
XX
PD   06-MAR-2001.
XX
PF   19-JAN-1998;   98US-0008697.
XX
PR   10-APR-1996;   96US-0011607.
PR   09-APR-1997;   97US-0835604.
XX
PA   (CHOW/) CHOW K L.
XX
PI   Chow KL;
XX
DR   WPI; 2001-256362/26.
DR   P-PSDB; AAB74419.
XX
PT   Detecting expression of mab-21 gene of Caenorhabditis elegans encoding
PT   novel protein required for choice of alternate cell fates in sample, by
PT   contacting mRNA from sample with nucleic acid hybridizing with mab-21 -
XX
PS   Disclosure; column 41-44; 79pp; English.
XX
CC   The present invention describes a method of detecting mab-21 expression
CC   in a sample involving contacting total mRNA from the sample with a probe
CC   to the mab-21 gene. The sequences of the mab-21 coding sequence and
CC   protein from several species are also given. The human mab-21 gene is
CC   found on chromosome 13q13, a region which has been linked to Moebius
CC   syndrome. The mab-21 protein shows homology to pax-6, mutations in which
CC   are associated with aniridia and which is linked to peripheral neural
CC   tissue differentiation.
XX
SQ   Sequence 2585 BP; 751 A; 540 C; 559 G; 688 T; 47 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 2585;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatt 24
           ||||||||||||||
Db    1058 CAAGTCACAGAATT 1045


RESULT  28
AAQ11851/c
ID   AAQ11851 standard; DNA; 3083 BP.
XX
AC   AAQ11851;
XX
DT   31-JUL-1991  (first entry)
XX
DE   Glutamate receptor 3.
XX
KW   Glutamate receptor 3; probe; ligand; drug screening; ss.
XX
OS   Rattus rattus.
XX
FH   Key             Location/Qualifiers
FT   CDS             167..2830
FT                   /*tag=  a
FT                   /product= GR3
FT   sig_peptide     167..232
FT                   /*tag=  b
FT   mat_peptide     233..2830
FT                   /*tag=  c
XX
PN   WO9106648-A.
XX
PD   16-MAY-1991.
XX
PF   25-OCT-1990;   90WO-US06153.
XX
PR   27-OCT-1989;   89US-0428116.
XX
PA   (SALK ) SALK INST FOR BIOL STUD.
XX
PI   Heinemann SF,  Boulter JR,  Hollmann M,  Bettler B,  Jensen JE;
XX
DR   WPI; 1991-164197/22.
DR   P-PSDB; AAR11991.
XX
PT   Glutamate receptors - used to screen for functional ligands and
PT   identify and isolate further receptors
XX
PS   Disclosure; Fig 4; 109pp; English.
XX
CC   GluR3 cDNA was isolated from a rat forebrain cDNA using a
CC   low-stringency screening protocol and a radiolabelled fragment of the
CC   GluR1 cDNA as probe.  The cDNA is deposited (ATCC 68133).
CC   The gene and protein can be used in drug screening, to
CC   determine whether a substance is a functional ligand for the
CC   receptor by monitoring ion channel activity.
CC   See also AAQ11849-855.
XX
SQ   Sequence 3083 BP; 877 A; 679 C; 735 G; 792 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 12;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 cattgtcgcaagtc 16
           ||||||||||||||
Db    3034 CATTGTCGCAAGTC 3021


RESULT  29
AAI61089/c
ID   AAI61089 standard; cDNA; 5662 BP.
XX
AC   AAI61089;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide.SEQ ID NO 5078.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
```

XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41933.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 5078; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 5662 BP; 1631 A; 1126 C; 1247 G; 1658 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 5662;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcaagtcacagaat 23
|||||||||||||||
Db 4497 GCAAGTCACAGAAT 4484

RESULT 30
AAC76014/c
ID AAC76014 standard; cDNA; 5773 BP.
XX
AC AAC76014;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1569 polynucleotide sequence SEQ ID NO:3137.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB41805.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 2354-2357; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;

Query Match 56.0%; Score 14; DB 21; Length 5773;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcaagtcacagaat 23
|||||||||||||||
Db 4611 GCAAGTCACAGAAT 4598

RESULT 31
AAI59303/c
ID AAI59303 standard; cDNA; 5876 BP.
XX
AC AAI59303;
XX
DT 22-OCT-2001 (first entry)
XX

DE Human polynucleotide SEQ ID NO 1506.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM40147.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 1506; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 5876 BP; 1704 A; 1165 C; 1288 G; 1719 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 5876;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcaagtcacagaat 23
||||||||||||||
Db 4739 GCAAGTCACAGAAT 4726

RESULT 32
AAF21307
ID AAF21307 standard; DNA; 32351 BP.
XX
AC AAF21307;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2874.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure; Page 1295-1303; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 32351 BP; 8594 A; 7026 C; 7405 G; 9326 T; 0 other;

```
  Query Match             56.0%;  Score 14;  DB 21;  Length 32351;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatta 25
           ||||||||||||||
Db   10696 aagtcacagaatta 10709


RESULT  33
AAA35185
ID   AAA35185 standard; DNA; 32351 BP.
XX
AC   AAA35185;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide 2nd SEQ ID NO:59.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 1212-1219; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 32351 BP; 8594 A; 7027 C; 7405 G; 9325 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 32351;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatta 25
           ||||||||||||||
Db   10696 aagtcacagaatta 10709


RESULT  34
AAF21311
ID   AAF21311 standard; DNA; 40298 BP.
XX
AC   AAF21311;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2878.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 1305-1315; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
```

```
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC   fragments and antisense oligonucleotides used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 40298 BP; 10485 A; 9119 C; 9484 G; 11210 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 40298;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatta 25
           ||||||||||||||
Db  11521 aagtcacagaatta 11534


RESULT  35
AAA35189
ID   AAA35189 standard; DNA; 40298 BP.
XX
AC   AAA35189;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide 2nd SEQ ID NO:63.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT   cancers -
XX
PS   Disclosure; Page 1221-1231; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an
CC   antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC   targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC   asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of
CC   the ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680
CC   (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC   invention. N.B. Sequences given in the disclosure of the present
CC   invention do not match up with their corresponding SEQ ID NO: sequences
CC   given in the sequence listing.
XX
SQ   Sequence 40298 BP; 10485 A; 9121 C; 9482 G; 11210 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 40298;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatta 25
           ||||||||||||||
Db  11521 aagtcacagaatta 11534


RESULT  36
AAZ01425
ID   AAZ01425 standard; DNA; 1038602 BP.
XX
AC   AAZ01425;
XX
DT   07-OCT-1999  (first entry)
XX
DE   Complete genome sequence of Chlamydia trachomatis.
XX
KW   Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW   paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW   nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW   bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX
OS   Chlamydia trachomatis.
XX
PN   WO9928475-A2.
XX
PD   10-JUN-1999.
XX
PF   27-NOV-1998;   98WO-IB01939.
XX
PR   04-NOV-1998;   98US-0107077.
PR   28-NOV-1997;   97FR-0015041.
PR   17-DEC-1997;   97FR-0016034.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-371125/31.
XX
PT   Genome sequence of Chlamydia trachomatis
XX
PS   Claim 1; Page 373-656; 1755pp; English.
XX
```

CC The present sequence represents the complete genome of Chlamydia
CC trachomatis. Open reading frames (ORFs) of the genome encode
CC polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC be used to control growth of the microorganism. Chlamydia trachomatis is
CC responsible for a large number of diseases, e.g. eye diseases such as
CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC conjunctivitis; genital diseases such as nongonococcal uretritis,
CC epidymitis, cervicitis, salpingitis, perihepatitis, bartholinitis;
CC pneumopathy in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
CC treating these diseases.
XX
SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 56.0%; Score 14; DB 20; Length 1038602;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 attgtcgcaagtca 17
||||||||||||||
Db 736214 attgtcgcaagtca 736227

RESULT 37
AAT22576
ID AAT22576 standard; cDNA to mRNA; 87 BP.
XX
AC AAT22576;
XX
DT 01-OCT-1996 (first entry)
XX
DE Human gene signature HUMGS04192.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI; 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 1160; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 87 BP; 31 A; 25 C; 9 G; 21 T; 1 other;

Query Match 52.0%; Score 13; DB 16; Length 87;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 aagtcacagaatt 24
|||||||||||||
Db 16 aagtcacagaatt 28

RESULT 38
AAT20233/c
ID AAT20233 standard; cDNA to mRNA; 167 BP.
XX
AC AAT20233;
XX
DT 24-JUL-1996 (first entry)
XX
DE Human gene signature HUMGS01380.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
XX
OS Homo sapiens.
XX
PN WO9514772-A1.
XX
PD 01-JUN-1995.
XX
PF 11-NOV-1994; 94WO-JP01916.
XX
PR 12-NOV-1993; 93JP-0355504.
XX
PA (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
XX
PI Matsubara K, Okubo K;
XX
DR WPI; 1995-206931/27.
XX
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
XX
PS Claim 1; Page 592; 2245pp; Japanese.
XX
CC A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in AAT19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC untranslated sequence is unique to a particular mRNA species, almost
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC is constructed so as to reflect accurately the relative abundance of
CC different mRNAs in the particular tissue from which it was derived.
CC The appearance frequency of a given GS in a cDNA library can be
CC determined (esp. using primers and probes derived from the GS
CC sequences) as a means of diagnosing abnormal cell function or for
CC recognising different cell types.
XX
SQ Sequence 167 BP; 65 A; 23 C; 26 G; 44 T; 9 other;

```
  Query Match            52.0%;  Score 13;  DB 16;  Length 167;
  Best Local Similarity  100.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db     118 AGTCACAGAATTA 106


RESULT  39
AAC26213/c
ID   AAC26213 standard; cDNA; 187 BP.
XX
AC   AAC26213;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 30288.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 30288; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 187 BP; 59 A; 27 C; 30 G; 71 T; 0 other;


  Query Match            52.0%;  Score 13;  DB 21;  Length 187;
  Best Local Similarity  100.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db     163 AAGTCACAGAATT 151


RESULT  40
AAA50159/c
ID   AAA50159 standard; cDNA; 363 BP.
XX
AC   AAA50159;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Antibody 5H7 heavy chain variable region cDNA.
XX
KW   5H7; monoclonal antibody; apoptosis; programmed cell death;
KW   major histocompatibility complex class I; MHC; tumour; therapy;
KW   graft-mediated immunoprotection; single chain antibody; scFv; ss.
XX
OS   Mammalia.
XX
PN   WO200047713-A2.
XX
PD   17-AUG-2000.
XX
PF   08-FEB-2000; 2000WO-US03234.
XX
PR   09-FEB-1999;   99US-0119238.
XX
PA   (ARCH-) ARCH DEV CORP.
XX
PI   Woodle ES,  Van Seventer JM,  Kulkarni S,  Kranz D,  Holman P;
XX
DR   WPI; 2000-558210/51.
DR   P-PSDB; AAY95776.
XX
PT   New polynucleotides encoding single chain light and heavy variable
PT   regions of a 5H7 antibody, useful in methods for conferring programmed
PT   cell death and graft-mediated immune protection of cell, tissue and
PT   organs -
XX
PS   Claim 2; Page 36; 48pp; English.
XX
CC   The present sequence is that of cDNA coding for the heavy chain
CC   variable region (VH) (see AAY95776) of 5H7, an anti-human class I
CC   major histocompatibility complex (MHC) monoclonal antibody (MAb)
CC   which recognises a monomorphic determinant of the alpha-3 domain.
CC   MAb 5H7 has been shown to induce programmed cell death (PCD) in
CC   lymphoid tumours, peripheral blood mononuclear cells, and B6 mouse
CC   splenocytes transgenic for class I HLA-B27.  Polynucleotides
CC   encoding 5H7 VL and VH sequences were used in the construction of
CC   single chain variable immunoglobulin domains (scFv) of the 5H7
CC   antibody (see AAY95781).  These can be used to confer PCD properties
CC   to cells, especially T and B lymphocyte tumour cells.  Recombinant
CC   DNA molecules encoding 5H7 scFv are also used in methods for
CC   conferring graft-mediated immune protection of cell, tissue and
CC   transplanted organs.
XX
SQ   Sequence 363 BP; 88 A; 101 C; 84 G; 90 T; 0 other;


  Query Match            52.0%;  Score 13;  DB 21;  Length 363;
  Best Local Similarity  100.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db     262 AAGTCACAGAATT 250


RESULT  41
AAQ25667/c
ID   AAQ25667 standard; cDNA; 378 BP.
XX
AC   AAQ25667;
XX
DT   28-DEC-1992  (first entry)
XX
```

```
DE   Sequence of the anti-urokinase antibody kappa variable region (VK)
DE   cDNA.
XX
KW   Chimeric monoclonal antibody; anti-urokinase antibody; PCR;
KW   antithrombotic agent; myocardial infarction therapy; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     1..9
FT                   /*tag= a
FT   mat_peptide     10..378
FT                   /*tag= b
FT   misc_feature    1..27
FT                   /*tag= c
FT                   /label= VH01 primer
FT   misc_feature    18..32
FT                   /*tag= d
FT                   /label= 5'VH1 primer
FT   misc_feature    340..371
FT                   /*tag= e
FT                   /label= 3'mVH primer
FT   misc_feature    356..371
FT                   /*tag= f
FT                   /label= JH01 primer
XX
PN   EP491351-A.
XX
PD   24-JUN-1992.
XX
PF   17-DEC-1991;    91EP-0121591.
XX
PR   18-DEC-1990;    90JP-0413829.
PR   11-NOV-1991;    91JP-0294464.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Iwasa S,  Taka H,  Watanabe T,  Tada H;
XX
DR   WPI; 1992-209528/26.
DR   P-PSDB; AAR24722.
XX
PT   Chimeric monoclonal antibodies - contain anti-human fibrin
PT   antibody light and heavy chain variable and constant for treating
PT   thrombotic conditions e.g. myocardial infarction
XX
PS   Example; Figure 15; 87pp; English.
XX
CC   Poly(A)+ RNA was prepd. from mouse anti-urokinase
CC   antibody-producing hybridoma UK1-3 cells. Using this poly(A) RNA as
CC   a template, an anti-urokinase antibody VK cDNA was amplified with
CC   the mC-kappa primer as primer for first strand sysnthesis, and the
CC   3'mV-kappa and 5'mV-kappa primers as primers for the PCR. The
CC   amplified fragment was restriction digested and ligated into a
CC   restriction fragment of pTB1423 to give an anti-urokinase antibody
CC   VK cDNA contg. plasmid pTB1456. The sequence of this plasmid is
CC   given in AAQ25667. The cDNA is a functional VK gene. The sequence of
CC   the primer (mC-gamma-1) is given in AAQ25689.
XX
SQ   Sequence 378 BP; 92 A; 100 C; 91 G; 95 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 13;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db     271 AAGTCACAGAATT 259


RESULT  42
AAA74604/c
ID   AAA74604 standard; cDNA; 408 BP.
XX
AC   AAA74604;
XX
DT   08-JAN-2001  (first entry)
XX
DE   Murine 5B3 antibody heavy chain variable region cDNA.
XX
KW   Mouse; 5B3 antibody; IgG1; piezoelectric immunoassay;
KW   small molecule explosive detection; 2,4,6-trinitrotoluene; TNT; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..408
FT                   /*tag= a
FT                   /partial
FT                   /product= "5B3 antibody heavy chain variable region"
XX
PN   WO200043774-A2.
XX
PD   27-JUL-2000.
XX
PF   25-JAN-2000; 2000WO-IL00048.
XX
PR   25-JAN-1999;   99IL-0128212.
XX
PA   (YISS ) YISSUM RES & DEV CO.
PA   (YEDA ) YEDA RES & DEV CO LTD.
XX
PI   Willner I,  Eshhar Z;
XX
DR   WPI; 2000-524259/47.
DR   P-PSDB; AAB15672.
XX
PT   Apparatus for detecting small molecules, especially explosives
PT   comprises a piezoelectric crystal -
XX
PS   Disclosure; Fig 3A; 90pp; English.
XX
CC   The present sequence encodes the heavy chain variable region of 5B3
CC   antibody. It was obtained from total RNA extracted from a 5B3 hybridoma
CC   by RT-PCR. 5B3 is an IgG1 antibody derived from a TNP-KLH immunised
CC   mouse. Its binding to TNP or DNS antigen can be blocked by very low
CC   amounts of TNT and it can therefore be used in a method for detecting
CC   small assayed explosive molecules. Molecules are detected using a
CC   piezoelectric sensor. Piezoelectric immunoassaying in liquid phase
CC   allows stationary and flow analysis of an aqueous sample. The method
CC   is sufficiently sensitive for detection of low molecular weight
CC   molecules.
XX
SQ   Sequence 408 BP; 90 A; 108 C; 99 G; 111 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db     316 AAGTCACAGAATT 304


RESULT  43
AAC56783/c
ID   AAC56783 standard; DNA; 415 BP.
XX
AC   AAC56783;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #654.
XX
```

```
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Pages 513-514; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 415 BP; 124 A; 99 C; 88 G; 104 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 415;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 attgtcgcaagtc 16
           |||||||||||||
Db     121 ATTGTCGCAAGTC 109


RESULT  44
AAC09392/c
ID   AAC09392 standard; cDNA; 476 BP.
XX
AC   AAC09392;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 13467.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 13467; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 476 BP; 109 A; 72 C; 90 G; 200 T; 5 other;


  Query Match             52.0%;  Score 13;  DB 21;  Length 476;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaat 23
           |||||||||||||
Db    263 CAAGTCACAGAAT 251


RESULT  45
AAH09733
ID   AAH09733 standard; cDNA; 556 BP.
XX
AC   AAH09733;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:6568.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
```

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3; SEQ ID 6568; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 556 BP; 148 A; 144 C; 165 G; 89 T; 10 other;

Query Match 52.0%; Score 13; DB 22; Length 556;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db     135 gcaagtcacagaa 147
```

Search completed: December 26, 2001, 14:16:28
Job time: 6980 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 10:59:28 ; Search time 143.1 Seconds
(without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196F-1
Perfect score: 25
Sequence: 1 gacattgtcgcaagtcacagaatta 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 9

Total number of hits satisfying chosen parameters: 8663

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
c	5	17	68.0	17	2	US-07-944-259-2	Sequence 2, Appli
	6	15	60.0	1771	2	US-09-166-203-48	Sequence 48, Appl
	7	15	60.0	1771	4	US-09-377-309-48	Sequence 48, Appl
c	8	14	56.0	1050	6	5498529-7	Patent No. 5498529
c	9	14	56.0	1677	2	US-08-372-652-14	Sequence 14, Appl
c	10	14	56.0	1677	5	PCT-US95-16311-14	Sequence 14, Appl
c	11	14	56.0	1787	2	US-08-372-652-6	Sequence 6, Appli
c	12	14	56.0	1787	5	PCT-US95-16311-6	Sequence 6, Appli
c	13	14	56.0	2233	1	US-08-496-631-1	Sequence 1, Appli
c	14	14	56.0	2585	4	US-09-008-697A-7	Sequence 7, Appli
c	15	14	56.0	3083	1	US-07-718-575-5	Sequence 5, Appli
c	16	14	56.0	3083	1	US-08-481-206-5	Sequence 5, Appli
c	17	14	56.0	3083	2	US-08-486-269A-5	Sequence 5, Appli
c	18	13	52.0	1662	1	US-08-651-572-1	Sequence 1, Appli
c	19	13	52.0	1662	3	US-09-066-544-1	Sequence 1, Appli
c	20	13	52.0	1662	3	US-08-951-086-1	Sequence 1, Appli
	21	13	52.0	2091	1	US-08-765-081-4	Sequence 4, Appli
	22	13	52.0	2091	3	US-09-098-082-4	Sequence 4, Appli
	23	13	52.0	2091	5	PCT-US95-06994-4	Sequence 4, Appli
c	24	13	52.0	2214	2	US-08-484-993B-1	Sequence 1, Appli
c	25	13	52.0	2214	2	US-08-484-158B-1	Sequence 1, Appli
c	26	13	52.0	2214	2	US-08-484-596A-1	Sequence 1, Appli
c	27	13	52.0	2214	2	US-08-480-150A-1	Sequence 1, Appli
c	28	13	52.0	2214	3	US-08-458-731-1	Sequence 1, Appli
c	29	13	52.0	2214	3	US-08-149-223A-1	Sequence 1, Appli
c	30	13	52.0	2381	2	US-08-484-993B-9	Sequence 9, Appli
c	31	13	52.0	2381	2	US-08-484-158B-9	Sequence 9, Appli
c	32	13	52.0	2381	2	US-08-484-596A-9	Sequence 9, Appli
c	33	13	52.0	2381	2	US-08-480-150A-9	Sequence 9, Appli
c	34	13	52.0	2381	3	US-08-458-731-9	Sequence 9, Appli
c	35	13	52.0	2381	3	US-08-149-223A-9	Sequence 9, Appli
c	36	13	52.0	4500	2	US-08-743-637B-35	Sequence 35, Appl
c	37	13	52.0	4500	3	US-08-526-840B-35	Sequence 35, Appl
	38	13	52.0	8041	1	US-08-765-081-1	Sequence 1, Appli
	39	13	52.0	8041	3	US-09-098-082-1	Sequence 1, Appli
	40	13	52.0	8041	5	PCT-US95-06994-1	Sequence 1, Appli
c	41	13	52.0	72928	3	US-09-009-913-1	Sequence 1, Appli
c	42	13	52.0	80246	4	US-09-078-294-4	Sequence 4, Appli
c	43	13	52.0	80595	4	US-09-078-294-3	Sequence 3, Appli
c	44	12	48.0	24	3	US-09-014-065-15	Sequence 15, Appl
	45	12	48.0	255	5	PCT-US93-03077-7	Sequence 7, Appli
c	46	12	48.0	279	1	US-08-047-041A-17	Sequence 17, Appl
c	47	12	48.0	308	4	US-09-232-522-3	Sequence 3, Appli
c	48	12	48.0	311	1	US-08-047-041A-4	Sequence 4, Appli
c	49	12	48.0	311	2	US-08-795-006A-24	Sequence 24, Appl
c	50	12	48.0	311	4	US-09-184-073-24	Sequence 24, Appl
	51	12	48.0	318	1	US-08-248-474-51	Sequence 51, Appl
	52	12	48.0	318	3	US-08-756-849-51	Sequence 51, Appl
c	53	12	48.0	354	3	US-09-065-059-12	Sequence 12, Appl
c	54	12	48.0	357	3	US-08-767-128-17	Sequence 17, Appl
c	55	12	48.0	411	1	US-08-137-117D-30	Sequence 30, Appl
c	56	12	48.0	411	1	US-08-436-717-30	Sequence 30, Appl
c	57	12	48.0	420	4	US-09-214-095D-109	Sequence 109, App
c	58	12	48.0	426	3	US-08-554-840-28	Sequence 28, Appl
c	59	12	48.0	427	2	US-08-484-956-111	Sequence 111, App
c	60	12	48.0	427	2	US-08-757-653-111	Sequence 111, App
c	61	12	48.0	498	2	US-08-484-956-113	Sequence 113, App
c	62	12	48.0	498	2	US-08-757-653-113	Sequence 113, App
	63	12	48.0	500	4	US-08-642-274D-18	Sequence 18, Appl
	64	12	48.0	500	4	US-08-952-014C-18	Sequence 18, Appl
	65	12	48.0	519	2	US-08-581-528A-5	Sequence 5, Appli
	66	12	48.0	519	5	PCT-US94-07799-5	Sequence 5, Appli
c	67	12	48.0	601	2	US-08-484-956-97	Sequence 97, Appl
	68	12	48.0	601	2	US-08-484-956-98	Sequence 98, Appl
c	69	12	48.0	601	2	US-08-484-956-99	Sequence 99, Appl
	70	12	48.0	601	2	US-08-484-956-100	Sequence 100, App
c	71	12	48.0	601	2	US-08-484-956-101	Sequence 101, App
	72	12	48.0	601	2	US-08-484-956-102	Sequence 102, App
c	73	12	48.0	601	2	US-08-484-956-107	Sequence 107, App
	74	12	48.0	601	2	US-08-484-956-108	Sequence 108, App
c	75	12	48.0	601	2	US-08-484-956-109	Sequence 109, App
	76	12	48.0	601	2	US-08-484-956-110	Sequence 110, App
c	77	12	48.0	601	2	US-08-757-653-97	Sequence 97, Appl
	78	12	48.0	601	2	US-08-757-653-98	Sequence 98, Appl
c	79	12	48.0	601	2	US-08-757-653-99	Sequence 99, Appl
	80	12	48.0	601	2	US-08-757-653-100	Sequence 100, App
c	81	12	48.0	601	2	US-08-757-653-101	Sequence 101, App
	82	12	48.0	601	2	US-08-757-653-102	Sequence 102, App
c	83	12	48.0	601	2	US-08-757-653-107	Sequence 107, App
	84	12	48.0	601	2	US-08-757-653-108	Sequence 108, App
c	85	12	48.0	601	2	US-08-757-653-109	Sequence 109, App
	86	12	48.0	601	2	US-08-757-653-110	Sequence 110, App
c	87	12	48.0	633	2	US-08-886-280-6	Sequence 6, Appli
c	88	12	48.0	633	3	US-08-870-370-10	Sequence 10, Appl
c	89	12	48.0	740	4	US-09-112-437-10	Sequence 10, Appl
c	90	12	48.0	800	1	US-08-472-239-1	Sequence 1, Appli
	91	12	48.0	807	1	US-07-706-691G-60	Sequence 60, Appl
	92	12	48.0	807	1	US-07-706-691G-62	Sequence 62, Appl
	93	12	48.0	807	1	US-07-706-691G-65	Sequence 65, Appl
	94	12	48.0	807	1	US-07-706-691G-66	Sequence 66, Appl
	95	12	48.0	807	1	US-07-706-691G-68	Sequence 68, Appl
	96	12	48.0	807	1	US-07-706-691G-69	Sequence 69, Appl
	97	12	48.0	807	1	US-07-706-691G-70	Sequence 70, Appl
	98	12	48.0	807	1	US-07-706-691G-71	Sequence 71, Appl
	99	12	48.0	807	1	US-07-706-691G-72	Sequence 72, Appl
	100	12	48.0	807	1	US-07-706-691G-73	Sequence 73, Appl

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db      21 GACATTGTCGCAAGTCACAGAATTA 45


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db      21 GACATTGTCGCAAGTCACAGAATTA 45


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db     137 GACATTGTCGCAAGTCACAGAATTA 161


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-06;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagtcacagaatta 25
           |||||||||||||||||||||||||
Db     137 GACATTGTCGCAAGTCACAGAATTA 161


RESULT   5
US-07-944-259-2/c
; Sequence 2, Application US/07944259
; Patent No. 5885792
;  GENERAL INFORMATION:
;    APPLICANT:  Ifuku, Ohji
;    APPLICANT:  Haze, Shinitiro
;    APPLICANT:  Kishimoto, Jiro
;    APPLICANT:  Nakahama, Kazuo
;    TITLE OF INVENTION:  BIOTIN OPERON
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wegner, Cantor, Mueller & Player
;      STREET:  1233 20th Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036-8218
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/944,259
;      FILING DATE:  19920914
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cantor, Herbert I.
;      REGISTRATION NUMBER:  24,392
;      REFERENCE/DOCKET NUMBER:  P-450-23557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-887-0400
;      TELEFAX:  202-835-0605
;      TELEX:  440706
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
US-07-944-259-2


  Query Match             68.0%;  Score 17;  DB 2;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 0.27;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gtcgcaagtcacagaat 23
           |||||||||||||||||
Db      17 GTCGCAAGTCACAGAAT 1


RESULT   6
US-09-166-203-48
; Sequence 48, Application US/09166203A
; Patent No. 5968826
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Condon, Tom P.
;  APPLICANT: Cowsert, Lex M.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
;  FILE REFERENCE: ISPH-0323
;  CURRENT APPLICATION NUMBER: US/09/166,203A
;  CURRENT FILING DATE: 1998-10-05
;  NUMBER OF SEQ ID NOS: 60
; SEQ ID NO 48
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1193)..(1387)
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (1709)..(1771)
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: L20788 Genbank
;   DATABASE ENTRY DATE: 1996-04-18
US-09-166-203-48


  Query Match             60.0%;  Score 15;  DB 2;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 2.8;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT   7
US-09-377-309-48
; Sequence 48, Application US/09377309B
; Patent No. 6258790
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Condon, Tom P.
;  APPLICANT: Cowsert, Lex M.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
;  FILE REFERENCE: ISPH-0390
;  CURRENT APPLICATION NUMBER: US/09/377,309B
;  CURRENT FILING DATE: 1999-08-19
;  EARLIER APPLICATION NUMBER: 09/166,203
;  EARLIER FILING DATE: 1998-10-05
;  NUMBER OF SEQ ID NOS: 99
; SEQ ID NO 48
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1193)..(1387)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1709)..(1771)
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: L20788 Genbank
;   DATABASE ENTRY DATE: 1996-04-18
US-09-377-309-48


  Query Match             60.0%;  Score 15;  DB 4;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 2.8;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT   8
5498529-7/c
;Patent No. 5498529
;    APPLICANT: BERKA, THOMAS R.;FORNWALD, JAMES A.;GORNIAK,
;JOSELINA G.;ROSENBERG, MARTIN;STRICKLER, JAMES E.;TAYLOR, DEAN P.
;    TITLE OF INVENTION: PROTEIN PROTEASE INHIBITORS FROM
;STREPTOMYCES
;    NUMBER OF SEQUENCES: 9
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/162,506
;      FILING DATE: 03-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 894,167
;      FILING DATE: 03-JUN-1992
;      APPLICATION NUMBER: 346,119
;      FILING DATE: 16-FEB-1989
;      APPLICATION NUMBER: 897,245
;      FILING DATE: 18-AUG-1986
;SEQ ID NO:7:
;      LENGTH: 1050
5498529-7


  Query Match             56.0%;  Score 14;  DB 6;  Length 1050;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatta 25
           ||||||||||||||
Db     251 AAGTCACAGAATTA 238


RESULT   9
US-08-372-652-14/c
; Sequence 14, Application US/08372652
; Patent No. 5932699
;  GENERAL INFORMATION:
;    APPLICANT:  Moore, David
;    APPLICANT:  Seol, Wongi
;    APPLICANT:  Choi, Hueng-Sik
;    TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;    TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street, Suite 3100
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/372,652
;      FILING DATE:  13-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  00786/246001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/542-5070
;      TELEFAX:  617/542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1677 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
US-08-372-652-14


  Query Match             56.0%;  Score 14;  DB 2;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     630 TTGTCGCAAGTCAC 617


RESULT  10
PCT-US95-16311-14/c
; Sequence 14, Application PC/TUS9516311
;  GENERAL INFORMATION:
```

```
;     APPLICANT:  Moore, David
;     APPLICANT:  Seol, Wongi
;     APPLICANT:  Choi, Hueng-Sik
;     TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;     TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street, Suite 3100
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/16311
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/372,652
;       FILING DATE:  13-JAN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Paul T.
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  00786/246001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1677 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
PCT-US95-16311-14


  Query Match             56.0%;  Score 14;  DB 5;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     630 TTGTCGCAAGTCAC 617


RESULT  11
US-08-372-652-6/c
; Sequence 6, Application US/08372652
; Patent No. 5932699
;  GENERAL INFORMATION:
;     APPLICANT:  Moore, David
;     APPLICANT:  Seol, Wongi
;     APPLICANT:  Choi, Hueng-Sik
;     TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;     TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street, Suite 3100
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/372,652
;       FILING DATE:  13-JAN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Paul T.
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  00786/246001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1787 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-372-652-6


  Query Match             56.0%;  Score 14;  DB 2;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     740 TTGTCGCAAGTCAC 727


RESULT  12
PCT-US95-16311-6/c
; Sequence 6, Application PC/TUS9516311
;  GENERAL INFORMATION:
;     APPLICANT:  Moore, David
;     APPLICANT:  Seol, Wongi
;     APPLICANT:  Choi, Hueng-Sik
;     TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;     TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street, Suite 3100
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/16311
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/372,652
;       FILING DATE:  13-JAN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Paul T.
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  00786/246001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1787 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  DNA
PCT-US95-16311-6


  Query Match             56.0%;  Score 14;  DB 5;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db     740 TTGTCGCAAGTCAC 727


RESULT  13
US-08-496-631-1/c
; Sequence 1, Application US/08496631
; Patent No. 5728548
;  GENERAL INFORMATION:
;    APPLICANT:  Bowman, Michael
;    TITLE OF INVENTION:  STEROID RECEPTOR RRI
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/496,631
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5248
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2233 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  363..1778
US-08-496-631-1


  Query Match             56.0%;  Score 14;  DB 1;  Length 2233;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ttgtcgcaagtcac 18
           ||||||||||||||
Db    1055 TTGTCGCAAGTCAC 1042


RESULT  14
US-09-008-697A-7/c
; Sequence 7, Application US/09008697A
; Patent No. 6197504
;  GENERAL INFORMATION:
;    APPLICANT:  Chow, King Lau
;    TITLE OF INVENTION:  USES OF MAB-21
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/008,697A
;      FILING DATE:  January 19, 1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Chan, Albert Wai-Kit
;      REGISTRATION NUMBER:  36,479
;      REFERENCE/DOCKET NUMBER:  50752-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2585 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-008-697A-7


  Query Match             56.0%;  Score 14;  DB 4;  Length 2585;
  Best Local Similarity   100.0%;  Pred. No. 9.9;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatt 24
           ||||||||||||||
Db    1058 CAAGTCACAGAATT 1045


RESULT  15
US-07-718-575-5/c
; Sequence 5, Application US/07718575
; Patent No. 5202257
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann Ph.D., Stephen F.
;    APPLICANT:  Boulter Ph.D., James R.
;    APPLICANT:  Hollmann Ph.D., Michael NMN
;    APPLICANT:  Bettler Ph.D., Bernhard NMN
;    APPLICANT:  Jensen Ph.D., Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 So. Flower St., Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071-2921
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/718,575
;      FILING DATE:  19910813
;      CLASSIFICATION:  435
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter Ph.D., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  P31 8962
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619)535-8949
;      TELEX:  9103330318
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3083 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  167..2833
US-07-718-575-5


  Query Match             56.0%;  Score 14;  DB 1;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 9.8;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 cattgtcgcaagtc 16
           ||||||||||||||
Db    3034 CATTGTCGCAAGTC 3021


RESULT  16
US-08-481-206-5/c
; Sequence 5, Application US/08481206
; Patent No. 5739291
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann Ph.D., Stephen F.
;    APPLICANT:  Boulter Ph.D., James R.
;    APPLICANT:  Hollmann Ph.D., Michael NMN
;    APPLICANT:  Bettler Ph.D., Bernhard NMN
;    APPLICANT:  Jensen Ph.D., Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 So. Flower St., Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071-2921
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/481,206
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/013,767
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter Ph.D., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  P31 8962
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619)535-8949
;      TELEX:  9103330318
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3083 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  167..2833
US-08-481-206-5


  Query Match             56.0%;  Score 14;  DB 1;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 9.8;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 cattgtcgcaagtc 16
           ||||||||||||||
Db    3034 CATTGTCGCAAGTC 3021


RESULT  17
US-08-486-269A-5/c
; Sequence 5, Application US/08486269A
; Patent No. 5945509
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann, Stephen F.
;    APPLICANT:  Boulter, James R.
;    APPLICANT:  Hollmann, Michael
;    APPLICANT:  Bettler, Bernhard
;    APPLICANT:  Jensen, Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS
;    TITLE OF INVENTION:  AND METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gray Cary Ware & Freidenrich LLP
;      STREET:  4365 Executive Drive, Suite 1600
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows DEMONSTRATION Version 2.0D
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/486,269A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/013,767
;      FILING DATE:  04-FEB-1993
;      APPLICATION NUMBER:  07/718,575
;      FILING DATE:  21-JUN-1991
;      APPLICATION NUMBER:  PCT/US90/06153
;      FILING DATE:  25-OCT-1990
;      APPLICATION NUMBER:  07/428,116
;      FILING DATE:  27-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9986
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-677-1409
;      TELEFAX:  619-677-1465
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3083 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR3
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  167...2830
;      OTHER INFORMATION:
US-08-486-269A-5


  Query Match           56.0%;  Score 14;  DB 2;  Length 3083;
  Best Local Similarity  100.0%;  Pred. No. 9.8;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 cattgtcgcaagtc 16
          ||||||||||||||
Db   3034 CATTGTCGCAAGTC 3021


RESULT  18
US-08-651-572-1/c
; Sequence 1, Application US/08651572
; Patent No. 5789228
;  GENERAL INFORMATION:
;    APPLICANT:  Lam, D. et al.
;    TITLE OF INVENTION:  Endoglucanases
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,572
;      FILING DATE:  Herewith
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-48
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1662 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-08-651-572-1


  Query Match           52.0%;  Score 13;  DB 1;  Length 1662;
  Best Local Similarity  100.0%;  Pred. No. 37;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db     974 AAGTCACAGAATT 962


RESULT  19
US-09-066-544-1/c
; Sequence 1, Application US/09066544
; Patent No. 6001984
;  GENERAL INFORMATION:
;    APPLICANT:  Lam, D. et al.
;    TITLE OF INVENTION:  Endoglucanases
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/066,544
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,572
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-48
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1662 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-09-066-544-1


  Query Match           52.0%;  Score 13;  DB 3;  Length 1662;
  Best Local Similarity  100.0%;  Pred. No. 37;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db     974 AAGTCACAGAATT 962


RESULT  20
US-08-951-086-1/c
; Sequence 1, Application US/08951086
; Patent No. 6074867
;  GENERAL INFORMATION:
;    APPLICANT:  Lam, D. et al.
;    TITLE OF INVENTION:  Endoglucanases
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  3.5 INCH DISKETTE
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  WORD PERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/951,086
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/651,572
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Charles J. Herron
;       REGISTRATION NUMBER:  28,019
;       REFERENCE/DOCKET NUMBER:  331400-48
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-994-1700
;       TELEFAX:  201-994-1744
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1662 NUCLEOTIDES
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  SINGLE
;       TOPOLOGY:  LINEAR
;     MOLECULE TYPE:  DNA
US-08-951-086-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 1662;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db     974 AAGTCACAGAATT 962


RESULT  21
US-08-765-081-4
; Sequence 4, Application US/08765081
; Patent No. 5798260
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium
;      OPERATING SYSTEM:  MS-Window 3.1
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,081
;      FILING DATE:  March 26, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997
;      REFERENCE/DOCKET NUMBER:  CHOR-1-10286
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0718 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  Corresponds to SEQ ID NO:1,
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
US-08-765-081-4


  Query Match             52.0%;  Score 13;  DB 1;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db     692 AGTCACAGAATTA 704


RESULT  22
US-09-098-082-4
; Sequence 4, Application US/09098082
; Patent No. 6040421
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium II
;      OPERATING SYSTEM:  MS-Windows 95
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/098,082
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/765,081
;      FILING DATE:  March 26, 1997
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  CHOR-1-12402
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0735 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  Corresponds to SEQ ID NO:1,
```

```
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
US-09-098-082-4


  Query Match             52.0%;  Score 13;  DB 3;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db     692 AGTCACAGAATTA 704


RESULT  23
PCT-US95-06994-4
; Sequence 4, Application PC/TUS9506994
;  GENERAL INFORMATION:
;    APPLICANT:  Children's Hospital & Medical Center
;    APPLICANT:  University of Washington
;    APPLICANT:  Washington State University Research Foundation
;    APPLICANT:  TARR, PHILLIP I
;    APPLICANT:  BILGE, SIMA S
;    APPLICANT:  BESSER, THOMAS E
;    APPLICANT:  VARY JR, JAMES C
;    TITLE OF INVENTION:  ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
;      STREET:  SUITE 2800, 1420 FIFTH AVENUE
;      CITY:  SEATTLE
;      STATE:  WASHINGTON
;      COUNTRY:  USA
;      ZIP:  WA 98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  07-JUN-95
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  24-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BRODERICK, THOMAS F
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  CHOR-18591
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682 8100
;      TELEFAX:  (206) 224 0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  corresponds to SEQ ID NO:1,
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
PCT-US95-06994-4


  Query Match             52.0%;  Score 13;  DB 5;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db     692 AGTCACAGAATTA 704


RESULT  24
US-08-484-993B-1/c
; Sequence 1, Application US/08484993B
; Patent No. 5837497
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,993B
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
```

```
;       LOCATION:  12..119
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  120..2153
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2153
US-08-484-993B-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  25
US-08-484-158B-1/c
; Sequence 1, Application US/08484158B
; Patent No. 5976545
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Pharmaceutical Compositions for
;    TITLE OF INVENTION:  Immunocontraception
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,158B
;      FILING DATE:  07-JUNE-95
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  09-NOV-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  32794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-484-158B-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  26
US-08-484-596A-1/c
; Sequence 1, Application US/08484596A
; Patent No. 5981228
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,596A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  11-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-484-596A-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  27
US-08-480-150A-1/c
; Sequence 1, Application US/08480150A
; Patent No. 5989550
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,150A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/149,223
;      FILING DATE:  09-NOV-1993
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-480-150A-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  28
US-08-458-731-1/c
; Sequence 1, Application US/08458731
; Patent No. 6001599
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,731
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
```

```
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2214 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Sus scrofa
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  12..119
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  120..2153
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2153
US-08-458-731-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  29
US-08-149-223A-1/c
; Sequence 1, Application US/08149223A
; Patent No. 6027727
;  GENERAL INFORMATION:
;     APPLICANT:  Harris Ph.D., Jeffrey D.
;     APPLICANT:  Hsu, Kuang T.
;     APPLICANT:  Podolski, Joseph S.
;     TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;     NUMBER OF SEQUENCES:  59
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;       STREET:  6300 Sears Tower, 233 South Wacker Drive
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  United States of America
;       ZIP:  60606-6402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/149,223A
;       FILING DATE:  09-NOV-1993
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/012,990
;       FILING DATE:  29-JAN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/973,341
;       FILING DATE:  09-NOV-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clough, David W.
;       REGISTRATION NUMBER:  36,107
;       REFERENCE/DOCKET NUMBER:  31745
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6653
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2214 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Sus scrofa
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  12..119
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  120..2153
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2153
US-08-149-223A-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  30
US-08-484-993B-9/c
; Sequence 9, Application US/08484993B
; Patent No. 5837497
;  GENERAL INFORMATION:
;     APPLICANT:  Harris Ph.D., Jeffrey D.
;     APPLICANT:  Hsu, Kuang T.
;     APPLICANT:  Podolski, Joseph S.
;     TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;     NUMBER OF SEQUENCES:  59
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;       STREET:  6300 Sears Tower, 233 South Wacker Drive
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  United States of America
;       ZIP:  60606-6402
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/484,993B
;       FILING DATE:  09-NOV-1993
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/012,990
;       FILING DATE:  29-JAN-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/973,341
;       FILING DATE:  09-NOV-1992
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-484-993B-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  31
US-08-484-158B-9/c
; Sequence 9, Application US/08484158B
; Patent No. 5976545
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Pharmaceutical Compositions for
;    TITLE OF INVENTION:  Immunocontraception
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,158B
;      FILING DATE:  07-JUNE-95
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  09-NOV-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  32794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-484-158B-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  32
US-08-484-596A-9/c
; Sequence 9, Application US/08484596A
; Patent No. 5981228
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,596A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  11-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
```

```
;       NAME:  Clough, David W.
;       REGISTRATION NUMBER:  36,107
;       REFERENCE/DOCKET NUMBER:  31745
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6653
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2381 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Canis familiaris
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  206..2353
US-08-484-596A-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  33
US-08-480-150A-9/c
; Sequence 9, Application US/08480150A
; Patent No. 5989550
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,150A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/149,223
;      FILING DATE:  09-NOV-1993
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-480-150A-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  34
US-08-458-731-9/c
; Sequence 9, Application US/08458731
; Patent No. 6001599
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,731
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2381 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Canis familiaris
;       DEVELOPMENTAL STAGE:  Juvenile
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  206..2353
US-08-458-731-9


  Query Match             52.0%;  Score 13;  DB 3;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  35
US-08-149-223A-9/c
; Sequence 9, Application US/08149223A
; Patent No. 6027727
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/149,223A
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-149-223A-9


  Query Match             52.0%;  Score 13;  DB 3;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaa 22
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  36
US-08-743-637B-35/c
; Sequence 35, Application US/08743637B
; Patent No. 5994066
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  PICARD, Francois J.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIES-SPECIFIC AND UNIVERSAL DNA
;    TITLE OF INVENTION:  PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
;    TITLE OF INVENTION:  IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
;    TITLE OF INVENTION:  ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 EAST WISCONSIN AVENUE
;      CITY:  MILWAUKEE
;      STATE:  WISCONSIN
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/743,637B
;      FILING DATE:  04-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,840
;      FILING DATE:  11-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414)277-5591
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4500 base pairs
```

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Streptococcus pneumoniae
US-08-743-637B-35


  Query Match             52.0%;  Score 13;  DB 2;  Length 4500;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db    1911 AAGTCACAGAATT 1899


RESULT  37
US-08-526-840B-35/c
; Sequence 35, Application US/08526840B
; Patent No. 6001564
;  GENERAL INFORMATION:
;     APPLICANT:  BERGERON, Michel G.
;     APPLICANT:  OUELLETTE, Marc
;     APPLICANT:  ROY, Paul H.
;     TITLE OF INVENTION:  SPECIFIC AND UNIVERSAL PROBES AND
;     TITLE OF INVENTION:  AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
;     TITLE OF INVENTION:  COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
;     TITLE OF INVENTION:  FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
;     NUMBER OF SEQUENCES:  177
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  QUARLES & BRADY
;       STREET:  411 East Wisconsin Avenue
;       CITY:  Milwaukee
;       STATE:  Wisconsin
;       COUNTRY:  USA
;       ZIP:  53202-4497
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/526,840B
;       FILING DATE:  11-SEP-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/304,732
;       FILING DATE:  12-SEP-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BAKER, Jean C.
;       REGISTRATION NUMBER:  35,433
;       REFERENCE/DOCKET NUMBER:  850586.90012
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (414) 277-5000
;       TELEFAX:  (414) 277-5591
;   INFORMATION FOR SEQ ID NO:  35:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4500 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Streptococcus pneumoniae
US-08-526-840B-35


  Query Match             52.0%;  Score 13;  DB 3;  Length 4500;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db    1911 AAGTCACAGAATT 1899


RESULT  38
US-08-765-081-1
; Sequence 1, Application US/08765081
; Patent No. 5798260
;  GENERAL INFORMATION:
;     APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;     TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;     NUMBER OF SEQUENCES:  12
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;       STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;       CITY:  Seattle
;       STATE:  Washington
;       COUNTRY:  USA
;       ZIP:  98101-2347
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;       COMPUTER:  IBM PC compatible/Pentium
;       OPERATING SYSTEM:  MS-Window 3.1
;       SOFTWARE:  Word for Windows-6.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/765,081
;       FILING DATE:  March 26, 1997
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/06994
;       FILING DATE:  June 7, 1995
;       APPLICATION NUMBER:  US 08/265,714
;       FILING DATE:  June 24, 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Shelton, Dennis K.
;       REGISTRATION NUMBER:  26,997
;       REFERENCE/DOCKET NUMBER:  CHOR-1-10286
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  1-206-682-8100; 1-206-224-0718 (direct)
;       TELEFAX:  1-206-224-0779
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8041 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Escherichia coli O157:H7
;       STRAIN:  86-24 NALR
;     IMMEDIATE SOURCE:
;       CLONE:  pEAR
US-08-765-081-1


  Query Match             52.0%;  Score 13;  DB 1;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db    3727 AGTCACAGAATTA 3739


RESULT  39
US-09-098-082-1
; Sequence 1, Application US/09098082
; Patent No. 6040421
;  GENERAL INFORMATION:
;     APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;     TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
```

```
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium II
;      OPERATING SYSTEM:  MS-Windows 95
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/098,082
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/765,081
;      FILING DATE:  March 26, 1997
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  CHOR-1-12402
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0735 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157:H7
;      STRAIN:  86-24 NALR
;    IMMEDIATE SOURCE:
;      CLONE:  pEAR
US-09-098-082-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db    3727 AGTCACAGAATTA 3739


RESULT  40
PCT-US95-06994-1
; Sequence 1, Application PC/TUS9506994
;  GENERAL INFORMATION:
;    APPLICANT:  Children's Hospital & Medical Center
;    APPLICANT:  University of Washington
;    APPLICANT:  Washington State University Research Foundation
;    APPLICANT:  TARR, PHILLIP I
;    APPLICANT:  BILGE, SIMA S
;    APPLICANT:  BESSER, THOMAS E
;    APPLICANT:  VARY JR, JAMES C
;    TITLE OF INVENTION:  ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
;      STREET:  SUITE 2800, 1420 FIFTH AVENUE
;      CITY:  SEATTLE
;      STATE:  WASHINGTON
;      COUNTRY:  USA
;      ZIP:  WA 98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  07-JUN-95
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  24-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BRODERICK, THOMAS F
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  CHOR-18591
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682 8100
;      TELEFAX:  (206) 224 0779
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157:H7
;      STRAIN:  86-24 NALR
;    IMMEDIATE SOURCE:
;      CLONE:  pEAR
PCT-US95-06994-1


  Query Match             52.0%;  Score 13;  DB 5;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatta 25
           |||||||||||||
Db    3727 AGTCACAGAATTA 3739


RESULT  41
US-09-009-913-1/c
; Sequence 1, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;    APPLICANT:  AxyS Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  Asthma Related Genes
;    NUMBER OF SEQUENCES:  339
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sherwood, Pamela J
;       REGISTRATION NUMBER:  36,677
;       REFERENCE/DOCKET NUMBER:  SEQ-4P
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650-327-3231
;       TELEFAX:  650-327-3231
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  72928 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
US-09-009-913-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 72928;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db   26365 AAGTCACAGAATT 26353


RESULT  42
US-09-078-294-4/c
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4


  Query Match             52.0%;  Score 13;  DB 4;  Length 80246;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db    5797 AAGTCACAGAATT 5785


RESULT  43
US-09-078-294-3/c
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 80595
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3


  Query Match             52.0%;  Score 13;  DB 4;  Length 80595;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 aagtcacagaatt 24
           |||||||||||||
Db    6081 AAGTCACAGAATT 6069


RESULT  44
US-09-014-065-15/c
; Sequence 15, Application US/09014065
; Patent No. 6033854
; GENERAL INFORMATION:
;  APPLICANT: Kurnit, David M.
;  APPLICANT: Chiang, Pei-Wen
;  APPLICANT: Wang, Chang-Ning J.
;  TITLE OF INVENTION: METHOD FOR DETERMINING THE COPY NUMBER OF A NUCLEIC ACID SEQUE
;  FILE REFERENCE: 06498/004001
;  CURRENT APPLICATION NUMBER: US/09/014,065
;  CURRENT FILING DATE: 1998-01-27
;  EARLIER APPLICATION NUMBER: US 08/434,474
;  EARLIER FILING DATE: 1995-05-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 15
;   LENGTH: 24
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-014-065-15


  Query Match             48.0%;  Score 12;  DB 3;  Length 24;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacaga 21
           ||||||||||||
Db      12 GCAAGTCACAGA 1


RESULT  45
PCT-US93-03077-7
; Sequence 7, Application PC/TUS9303077
;  GENERAL INFORMATION:
;     APPLICANT:  Board of Regents, The Universityof Texas System
;     APPLICANT:  Gaynor, Richard B.
;     APPLICANT:  Wu, Foon Kin
;     TITLE OF INVENTION:  PROTEIN CELLULAR FACTOR USEFUL FOR
;     TITLE OF INVENTION:  REGULATING GENE EXPRESSION
;     NUMBER OF SEQUENCES:  7
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Arnold White & Durkee
;       STREET:  P.O. Box 4433
;       CITY:  Houston
;       STATE:  Texas
;       COUNTRY:  USA
;       ZIP:  77210
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/03077
;       FILING DATE:  19930331
```

```
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/862,025
;       FILING DATE:  April 2, 1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Kammerer, Patricia A.
;       REGISTRATION NUMBER:  29,775
;       REFERENCE/DOCKET NUMBER:  UTFD270PCT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  713-787-1540
;       TELEFAX:  713-749-2679
;       TELEX:
;  INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  255 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  unknown
;       TOPOLOGY:  unknown
PCT-US93-03077-7


  Query Match             48.0%;  Score 12;  DB 5;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   12;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 agtcacagaatt 24
           ||||||||||||
Db     206 AGTCACAGAATT 217


Search completed: December 26, 2001, 12:52:29
Job time: 6781 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:12:33 ; Search time 4619.78 Seconds
(without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196F-1
Perfect score: 25
Sequence: 1 gacattgtcgcaagtcacagaatta 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 9

Total number of hits satisfying chosen parameters: 531357

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	17	68.0	357	11	BE800706	BE800706 sq97c06.y
	2	17	68.0	388	10	AW781150	AW781150 s189g06.y
c	3	17	68.0	751	10	BE658717	BE658717 GM700007A
c	4	17	68.0	1318	11	BG871816	BG871816 602792574
	5	16	64.0	299	10	BB450863	BB450863 BB450863
c	6	16	64.0	418	11	BF556014	BF556014 UI-R-A1-d
	7	15	60.0	172	11	BG977588	BG977588 PM4-CI008
c	8	15	60.0	214	10	AW871559	AW871559 ra79d04.y
	9	15	60.0	233	10	BE763893	BE763893 RC4-NT005
	10	15	60.0	253	10	BE763892	BE763892 RC4-NT005
	11	15	60.0	256	10	AA317110	AA317110 EST19001
c	12	15	60.0	261	11	Z45844	Z45844 HSCZVE041 n
c	13	15	60.0	263	10	BB141316	BB141316 BB141316
	14	15	60.0	263	11	BF745516	BF745516 CM2-BT082
c	15	15	60.0	282	11	BF800965	BF800965 QV4-CI010
	16	15	60.0	286	10	AA344205	AA344205 EST50086
c	17	15	60.0	292	13	B53227	B53227 CIT-HSP-200
c	18	15	60.0	317	11	BF743951	BF743951 CM2-BT082
c	19	15	60.0	324	13	BH121038	BH121038 RPCI-24-2
c	20	15	60.0	327	11	T30901	T30901 EST24536 Hu
c	21	15	60.0	335	11	BF743964	BF743964 CM2-BT082
c	22	15	60.0	346	11	BF743954	BF743954 CM2-BT082
c	23	15	60.0	349	11	BG466199	BG466199 EtESTed25
	24	15	60.0	353	10	AA648449	AA648449 ns22f10.s
	25	15	60.0	370	11	T07051	T07051 EST04940 Fe
	26	15	60.0	402	10	AI694429	AI694429 wd83f06.x
	27	15	60.0	404	10	AI203847	AI203847 qf75h03.x
	28	15	60.0	405	11	BF726110	BF726110 by01c01.x
c	29	15	60.0	406	11	BG413230	BG413230 EtESTed18
	30	15	60.0	416	10	AI942464	AI942464 wp80c05.x
	31	15	60.0	420	10	AA172188	AA172188 zp29f06.s
c	32	15	60.0	427	13	AQ142435	AQ142435 HS_3064_A
c	33	15	60.0	429	10	AW870882	AW870882 ra51h09.y
	34	15	60.0	454	13	AQ086771	AQ086771 HS_2175_B
	35	15	60.0	455	10	AA973493	AA973493 oo45e05.s
c	36	15	60.0	456	10	AI756736	AI756736 EtESTea24
	37	15	60.0	457	10	AI361386	AI361386 qy47f01.x
c	38	15	60.0	464	11	BG562271	BG562271 EtESTee05
	39	15	60.0	471	10	AI635259	AI635259 tz79h07.x
	40	15	60.0	485	11	BG020076	BG020076 dc47c07.x
c	41	15	60.0	487	11	BG516037	BG516037 EtESTed48
c	42	15	60.0	489	11	BF775935	BF775935 286217 MA
	43	15	60.0	498	13	AZ398330	AZ398330 1M0163N15
	44	15	60.0	502	10	AA846453	AA846453 aj56b03.s
	45	15	60.0	502	10	AI672463	AI672463 wa03e05.x
	46	15	60.0	503	10	AI634722	AI634722 tz09e04.x
	47	15	60.0	512	10	AW418528	AW418528 xr32c07.x
	48	15	60.0	512	13	AQ801291	AQ801291 HS_2088_A
	49	15	60.0	520	13	AQ602419	AQ602419 HS_2128_A
	50	15	60.0	531	10	BE043069	BE043069 ho32e02.x
c	51	15	60.0	531	13	AQ806020	AQ806020 HS_3237_A
c	52	15	60.0	552	10	AL375298	AL375298 MtBB12F10
	53	15	60.0	554	10	AI268404	AI268404 qm05e10.x
c	54	15	60.0	562	13	AQ483592	AQ483592 RPCI-11-2
c	55	15	60.0	563	13	AQ430099	AQ430099 HS_5074_A
c	56	15	60.0	587	13	AZ554645	AZ554645 RPCI-23-1
c	57	15	60.0	601	10	AW828925	AW828925 ra69g07.y
	58	15	60.0	617	10	AW316985	AW316985 xx10c10.x
	59	15	60.0	626	10	AV657973	AV657973 AV657973
c	60	15	60.0	652	13	AQ541562	AQ541562 RPCI-11-3
c	61	15	60.0	657	11	BF726111	BF726111 by01c01.y
c	62	15	60.0	658	13	AZ816912	AZ816912 2M0085C19
c	63	15	60.0	665	13	AZ459164	AZ459164 1M0263K18
	64	15	60.0	691	13	AZ963904	AZ963904 2M0233D23
	65	15	60.0	912	10	AA626775	AA626775 ad09e07.s
	66	15	60.0	1146	11	BG174876	BG174876 602336157
	67	15	60.0	1239	10	BE735021	BE735021 601569755
	68	14	56.0	61	10	AI174176	AI174176 uh98f03.r
c	69	14	56.0	114	11	BF987752	BF987752 CM0-GN016
c	70	14	56.0	116	11	BF991166	BF991166 CM0-GN016
	71	14	56.0	132	13	AZ073992	AZ073992 RPCI-23-4
c	72	14	56.0	164	10	BE090319	BE090319 RC3-BT071
c	73	14	56.0	182	11	BI416828	BI416828 hasp002xg
	74	14	56.0	182	13	AZ074961	AZ074961 RPCI-23-4
	75	14	56.0	190	11	F07940	F07940 HSC2MA061 n
c	76	14	56.0	192	13	AZ637253	AZ637253 1M0496I02
c	77	14	56.0	193	13	AQ542941	AQ542941 RPCI-11-3
	78	14	56.0	197	10	AA957877	AA957877 UI-R-E1-f
	79	14	56.0	201	10	AV530463	AV530463 AV530463
c	80	14	56.0	201	10	BE684330	BE684330 184699 MA
c	81	14	56.0	204	10	AW579956	AW579956 PM2-HT035
c	82	14	56.0	205	11	BF926840	BF926840 IL5-NT022
c	83	14	56.0	226	10	AV345479	AV345479 AV345479
	84	14	56.0	255	11	BF353051	BF353051 IL3-HT061
	85	14	56.0	255	13	AZ559921	AZ559921 RPCI-23-2

	86	14	56.0	259	10	AA766388 AA766388 oa36c12.s
c	87	14	56.0	264	10	AA329990 AA329990 EST33660
	88	14	56.0	267	13	B41289 B41289 HS-1053-B1-
	89	14	56.0	268	11	L44566 L44566 HUMEST6D8 H
c	90	14	56.0	269	10	AA984664 AA984664 am87h02.s
c	91	14	56.0	270	10	BE181705 BE181705 QV1-HT068
c	92	14	56.0	270	13	AQ532982 AQ532982 RPCI-11-3
c	93	14	56.0	276	13	AQ414837 AQ414837 RPCI-11-1
c	94	14	56.0	276	13	AQ471115 AQ471115 CITBI-E1-
c	95	14	56.0	282	11	BG049539 BG049539 EM1_5_D06
c	96	14	56.0	287	13	AQ415921 AQ415921 RPCI-11-1
c	97	14	56.0	289	13	AQ070825 AQ070825 HS_2256_A
c	98	14	56.0	290	11	BF804635 BF804635 RC2-CI008
	99	14	56.0	290	13	AQ634826 AQ634826 RPCI-11-4
	100	14	56.0	301	13	AQ352850 AQ352850 CITBI-E1-

ALIGNMENTS

RESULT 1
BE800706
LOCUS BE800706 357 bp mRNA EST 20-SEP-2000
DEFINITION sq97c06.y1 Gm-c1049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1049-1091 5' similar to TR:O82150 O82150 CHLOROPLAST FTSH PROTEASE. ;, mRNA sequence.
ACCESSION BE800706
VERSION BE800706.1 GI:10231818
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 357)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Putative full length read
vector to vector length is 364
High quality sequence stop: 340.
FEATURES Location/Qualifiers
source 1. .357
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1049-1091"
/clone_lib="Gm-c1049"
/tissue_type="whole seedlings of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The Clark NIL was constructed and seed was provided by Dr. J. Specht, University of Nebraska (Shoemaker and Specht, 1995). The cDNA library was constructed from mRNA isolated from whole seedlings of 3 week old greenhouse grown plants. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site and a 3' anchor. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). The library was constructed in cooperation with Dr. Paul Keim's laboratory at Northern Arizona University."
BASE COUNT 169 a 53 c 47 g 88 t
ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 357;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 cgcaagtcacagaatta 25
||||||||||||||||||
Db 138 CGCAAGTCACAGAATTA 154

RESULT 2
AW781150
LOCUS AW781150 388 bp mRNA EST 12-MAY-2000
DEFINITION sl89g06.y1 Gm-c1037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1037-1235 5' similar to TR:O82150 O82150 CHLOROPLAST FTSH PROTEASE. ;, mRNA sequence.
ACCESSION AW781150
VERSION AW781150.1 GI:7795753
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 388)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 375.
FEATURES Location/Qualifiers
source 1. .388
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1037-1235"
/clone_lib="Gm-c1037"
/tissue_type="fully expanded leaves of greenhouse grown plants"
/dev_stage="2 week old"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This cDNA library was constructed from mRNA isolated from fully expanded leaves of greenhouse grown plants that were 2

weeks old. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restrictions site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electro- Max DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana- Champaign. email: l-vodkin@uiuc.edu"

BASE COUNT 123 a 67 c 92 g 106 t
ORIGIN

Query Match 68.0%; Score 17; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 cgcaagtcacagaatta 25
||||||||||||||||||
Db 334 CGCAAGTCACAGAATTA 350

RESULT 3
BE658717/c
LOCUS BE658717 751 bp mRNA EST 24-MAY-2001
DEFINITION GM700007A10F9 Gm-r1070 Glycine max cDNA clone Gm-r1070-2561 3', mRNA sequence.
ACCESSION BE658717
VERSION BE658717.1 GI:9984609
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 751)
AUTHORS Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
TITLE A Functional Genomics Program for Soybean (NSF 9872565)
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: AW156684 corresponding to Gm-c1015-2632 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact:clones@genomesystems.com or info@genome ystems.com web site:www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES Location/Qualifiers
source 1. .751
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="Gm-r1070-2561"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, reracked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were reracked to form library Gm-r1070. The cDNA clones of the reracked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html . Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT 193 a 171 c 131 g 226 t 30 others
ORIGIN

Query Match 68.0%; Score 17; DB 10; Length 751;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 cgcaagtcacagaatta 25
||||||||||||||||||
Db 112 CGCAAGTCACAGAATTA 96

RESULT 4
BG871816/c
LOCUS BG871816 1318 bp mRNA EST 29-MAY-2001
DEFINITION 602792574F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4923619 5', mRNA sequence.
ACCESSION BG871816
VERSION BG871816.1 GI:14222356
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1318)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10845 row: d column: 20
High quality sequence start: 3
High quality sequence stop: 211.
FEATURES Location/Qualifiers
source 1. .1318
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4923619"
/clone_lib="NCI_CGAP_SG2"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 379 a 339 c 297 g 303 t
ORIGIN

```
  Query Match             68.0%;  Score 17;  DB 11;  Length 1318;
  Best Local Similarity   100.0%;  Pred. No. 9.2;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 cattgtcgcaagtcaca 19
           |||||||||||||||||
Db    1085 CATTGTCGCAAGTCACA 1069


RESULT   5
BB450863
LOCUS       BB450863      299 bp    mRNA            EST       21-JUL-2000
DEFINITION  BB450863 RIKEN full-length enriched, 12 days embryo spinal ganglion
            Mus musculus cDNA clone D130006I17 3', mRNA sequence.
ACCESSION   BB450863
VERSION     BB450863.1  GI:9311898
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 299)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .299
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D130006I17"
                     /clone_lib="RIKEN full-length enriched, 12 days embryo
                     spinal ganglion"
                     /tissue_type="spinal ganglion"
                     /dev_stage="12 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       75 a     72 c     45 g    107 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 299;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 cgcaagtcacagaatt 24
           ||||||||||||||||
Db       3 CGCAAGTCACAGAATT 18


RESULT   6
BF556014/c
LOCUS       BF556014      418 bp    mRNA            EST       12-DEC-2000
DEFINITION  UI-R-A1-dz-d-02-0-UI.r1 UI-R-A1 Rattus norvegicus cDNA clone
            UI-R-A1-dz-d-02-0-UI 5', mRNA sequence.
ACCESSION   BF556014
VERSION     BF556014.1  GI:11665744
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1771157
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-A1-dz-d-02-0-UI"
                     /clone_lib="UI-R-A1"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-A1
                     library is a subtracted library derived from the UI-R-A0
                     library. The UI-R-A0 library  consisted of a mixture of
                     individually tagged normalized libraries constructed from
                     rat placenta, adult lung, brain, liver, kidney, heart,
```

spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not 1 site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-A1) was constructed as follows: PCR amplified cDNA inserts from a pool of approximately 3,840 UI-R-A0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-A0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-A1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

BASE COUNT 121 a 97 c 117 g 83 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 418;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 gacattgtcgcaagtc 16
          ||||||||||||||||
Db    110 GACATTGTCGCAAGTC 95
```

RESULT 7
BG977588
LOCUS BG977588 172 bp mRNA EST 12-JUN-2001
DEFINITION PM4-CI0084-040101-006-d11 CI0084 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG977588
VERSION BG977588.1 GI:14380323
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM4&t2=PM4-CI0084-040101-006-d11&t3=2001-01-04&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 165.

FEATURES Location/Qualifiers
source 1. .172
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CI0084"
/dev_stage="Adult"
/note="Organ: colon_ins; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 37 a 51 c 55 g 28 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 172;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      7 gtcgcaagtcacaga 21
          |||||||||||||||
Db      8 GTCGCAAGTCACAGA 22
```

RESULT 8
AW871559/c
LOCUS AW871559 214 bp mRNA EST 10-MAY-2001
DEFINITION ra79d04.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita cDNA 5' similar to WP:C26D10.4 CE03027 ;, mRNA sequence.
ACCESSION AW871559
VERSION AW871559.1 GI:8005612
KEYWORDS EST.
SOURCE southern root-knot nematode.
ORGANISM Meloidogyne incognita
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina; Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE 1 (bases 1 to 214)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Uma Rao and David Bird (david_bird@ncsu.edu) at North Carolina State University. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .214
/organism="Meloidogyne incognita"
/db_xref="taxon:6306"
/clone_lib="Bird-Rao Meloidogyne incognita J2"
/dev_stage="enriched for 2nd stage juveniles"
/lab_host="XLORL"
/note="Vector: ZAP express - pBKCMV (Stratagene); Site_1: EcoRI; Site_2: XhoI; Oligo (dT) primed library. cDNA was contructed and cloned unidirectionally into the vector within the 5' EcoRI and 3' XhoI sites. This library was constructed by Dr. Uma Rao and Dr. David Bird at North Carolina state University."

BASE COUNT 59 a 52 c 19 g 84 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 214;

```
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db     184 CAAGTCACAGAATTA 170


RESULT   9
BE763893
LOCUS       BE763893      233 bp    mRNA            EST       19-SEP-2000
DEFINITION  RC4-NT0054-120600-015-g05 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE763893
VERSION     BE763893.1  GI:10193817
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=RC4-NT0054-120
            600-015-g05&t3=2000-06-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 233.
FEATURES             Location/Qualifiers
     source          1. .233
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0054"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       88 a     27 c     52 g     66 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 233;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db      94 CAAGTCACAGAATTA 108


RESULT  10
BE763892
LOCUS       BE763892      253 bp    mRNA            EST       19-SEP-2000
DEFINITION  RC4-NT0054-120600-015-g03 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE763892
VERSION     BE763892.1  GI:10193816
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 253)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=RC4-NT0054-120
            600-015-g03&t3=2000-06-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 12
            High quality sequence stop: 253.
FEATURES             Location/Qualifiers
     source          1. .253
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0054"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       95 a     31 c     56 g     71 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 253;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db     114 CAAGTCACAGAATTA 128


RESULT  11
AA317110
LOCUS       AA317110      256 bp    mRNA            EST       19-APR-1997
DEFINITION  EST19001 Lung Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA317110
VERSION     AA317110.1  GI:1969448
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 256)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
```

,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W. , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.

TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)

MEDLINE 96026280

COMMENT Other_ESTs: THC99418
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES Location/Qualifiers
source 1. .256
/organism="Homo sapiens"
/db_xref="ATCC (inhost):117571"
/db_xref="taxon:9606"
/clone_lib="Lung"
/dev_stage="adult"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI ; Site_2: XhoI"

BASE COUNT 73 a 57 c 48 g 75 t 3 others

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 256;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     189 CAAGTCACAGAATTA 203
```

RESULT 12
Z45844/c
LOCUS Z45844 261 bp mRNA EST 14-NOV-1994
DEFINITION HSCZVE041 normalized infant brain cDNA Homo sapiens cDNA clone c-zve04, mRNA sequence.
ACCESSION Z45844
VERSION Z45844.1 GI:575078
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 261)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y., Sebastiani-Kabaktchis,C. and Tessier,A.

TITLE IMAGE: molecular integration of the analysis of the human genome and its expression

JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

MEDLINE 95277534

COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33 1 6[illegible]172800
Fax: 33 1 6[illegible]778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-zve04
Seq primer: (-21)M13_universal.

FEATURES Location/Qualifiers
source 1. .261
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-zve04"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII; Site_2: NotI; sex=Female; dev_stage=3 months old; isolate=muscular atrophy patient; tissue_type=total brain ; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Soares, Psychiatry Dept. Columbia University, USA. Normalization_method: Bento Soares, P.N.A.S in press"

BASE COUNT 77 a 56 c 31 g 97 t

ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     138 CAAGTCACAGAATTA 124
```

RESULT 13
BB141316/c
LOCUS BB141316 263 bp mRNA EST 28-JUN-2000
DEFINITION BB141316 RIKEN full-length enriched, adult female vagina Mus musculus cDNA clone 9930007O04 3', mRNA sequence.
ACCESSION BB141316
VERSION BB141316.1 GI:8796253
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 263)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Konno,H., et al.)

JOURNAL Unpublished (2000)

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .263
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="9930007O04"
/clone_lib="RIKEN full-length enriched, adult female vagina"
/sex="female"
/tissue_type="vagina"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 69 a 56 c 43 g 95 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
||||||||||||||||
Db 62 CAAGTCACAGAATTA 48

RESULT 14
BF745516
LOCUS BF745516 263 bp mRNA EST 10-JAN-2001
DEFINITION CM2-BT0826-021100-469-h05 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF745516
VERSION BF745516.1 GI:12072192
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 263)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-021100-469-h05&t3=2000-11-02&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 129
High quality sequence stop: 263.

FEATURES Location/Qualifiers
source 1. .263
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0826"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 ,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 77 a 55 c 60 g 71 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 gcaagtcacagaatt 24
|||||||||||||||
Db 217 GCAAGTCACAGAATT 231

RESULT 15
BF800965/c
LOCUS BF800965 282 bp mRNA EST 12-JAN-2001
DEFINITION QV4-CI0100-231000-457-d12 CI0100 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF800965
VERSION BF800965.1 GI:12129954
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 282)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663

```
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV4&t2=QV4-CI0100-
            231000-457-d12&t3=2000-10-23&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 59
            High quality sequence stop: 282.
FEATURES             Location/Qualifiers
     source          1. .282
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CI0100"
                     /dev_stage="Adult"
                     /note="Organ: colon_ins; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       68 a     75 c     67 g     72 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 282;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 attgtcgcaagtcac 18
          ||||||||||||||||
Db     51 ATTGTCGCAAGTCAC 37


RESULT  16
AA344205
LOCUS       AA344205       286 bp    mRNA            EST       21-APR-1997
DEFINITION  EST50086 Gall bladder I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA344205
VERSION     AA344205.1  GI:1996443
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC99418
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .286
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):145738"
                     /db_xref="taxon:9606"
                     /clone_lib="Gall bladder I"
                     /sex="female"
                     /dev_stage="adult, 25 yrs"
                     /note="Organ: gall bladder; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       83 a     62 c     48 g     91 t      2 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 286;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     11 caagtcacagaatta 25
          ||||||||||||||||
Db    197 CAAGTCACAGAATTA 211


RESULT  17
B53227/c
LOCUS       B53227         292 bp    DNA             GSS       20-JUN-1998
DEFINITION  CIT-HSP-2009H2.TR CIT-HSP Homo sapiens genomic clone 2009H2, DNA
            sequence.
ACCESSION   B53227
VERSION     B53227.1  GI:2607561
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
            and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Homo sapiens"
                     /db_xref="GDB:7040812"
                     /db_xref="taxon:9606"
                     /clone="2009H2"
```

```
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT       58 a     64 c     90 g     80 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     248 CAAGTCACAGAATTA 234


RESULT  18
BF743951/c
LOCUS       BF743951       317 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-c02 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743951
VERSION     BF743951.1  GI:12070627
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-BT0826-
            181000-428-c02&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 17
            High quality sequence stop: 317.
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0826"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       96 a     71 c     49 g    101 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db      45 GCAAGTCACAGAATT 31


RESULT  19
BH121038/c
LOCUS       BH121038       324 bp    DNA             GSS       19-JUL-2001
DEFINITION  RPCI-24-230E13.TV RPCI-24 Mus musculus genomic clone RPCI-24-230E13
            , DNA sequence.
ACCESSION   BH121038
VERSION     BH121038.1  GI:14964550
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-230E13.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 230  row: E  column: 13
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-230E13"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT       93 a     57 c     88 g     86 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 324;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db      26 GCAAGTCACAGAATT 12


RESULT  20
T30901/c
LOCUS       T30901         327 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST24536 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA
```

```
            sequence.
ACCESSION   T30901
VERSION     T30901.1  GI:612999
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W., Clayton,R.A.,
            Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., FitzGerald
            ,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S., Kelley,J.M.,
            Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J.,
            Dimke,D., Feng,P., Ferrie,A., Fischer,C., Hastings,G.A., He,W.-W.,
            Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Kozak,D.L.,
            Kunsch,C., Ji,H., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei
            ,Y.-F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC10632
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .327
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):100882"
                      /db_xref="taxon:9606"
                      /clone_lib="Human Brain"
                      /note="Organ: brain"
BASE COUNT       89 a     75 c     47 g    115 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||||
Db     295 CAAGTCACAGAATTA 281


RESULT  21
BF743964/c
LOCUS       BF743964      335 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-h04 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743964
VERSION     BF743964.1  GI:12070640
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-BT0826-
            181000-428-h04&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 40
            High quality sequence stop: 335.
FEATURES              Location/Qualifiers
     source           1. .335
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="BT0826"
                      /dev_stage="Adult"
                      /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT      101 a     76 c     55 g    103 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 335;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db      61 GCAAGTCACAGAATT 47


RESULT  22
BF743954/c
LOCUS       BF743954      346 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-d07 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743954
VERSION     BF743954.1  GI:12070630
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 346)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
```

```
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-BT0826-
            181000-428-d07&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 43
            High quality sequence stop: 346.
FEATURES             Location/Qualifiers
     source          1. .346
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0826"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      103 a     78 c     54 g    111 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 346;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db      73 GCAAGTCACAGAATT 59


RESULT  23
BG466199/c
LOCUS       BG466199      349 bp    mRNA            EST       20-MAR-2001
DEFINITION  EtESTed25c05.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5' similar to TR:O88349 O88349 LATENT TGF BETA BINDING
            PROTEIN. ;, mRNA sequence.
ACCESSION   BG466199
VERSION     BG466199.1  GI:13395174
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 318.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT      75 a     107 c    103 g     64 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 349;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagt 15
           |||||||||||||||
Db     328 GACATTGTCGCAAGT 314


RESULT  24
AA648449
LOCUS       AA648449      353 bp    mRNA            EST       13-NOV-1997
DEFINITION  ns22f10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1184395 3'
            similar to contains element LTR5 repetitive element ;, mRNA
            sequence.
ACCESSION   AA648449
VERSION     AA648449.1  GI:2574878
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            unknown library type
            Insert Length: 2027   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 334.
FEATURES             Location/Qualifiers
     source          1. .353
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1184395"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
```

primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 131 a 49 c 39 g 134 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 353;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
||||||||||||||||
Db 249 CAAGTCACAGAATTA 263

RESULT 25
T07051
LOCUS T07051 370 bp mRNA EST 30-JUN-1993
DEFINITION EST04940 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBEC79, mRNA sequence.
ACCESSION T07051
VERSION T07051.1 GI:318200
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 370)
AUTHORS Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
TITLE 3,400 expressed sequence tags identify diversity of transcripts from human brain
JOURNAL Nature Genet. 4, 256-267 (1993)
MEDLINE 93364420
COMMENT Contact: Adams, MD
The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13-21.
FEATURES Location/Qualifiers
source 1. .370
/organism="Homo sapiens"
/db_xref="ATCC (inhost):83698"
/db_xref="taxon:9606"
/clone="HFBEC79"
/clone_lib="Fetal brain, Stratagene (cat#936206)"
/note="Vector: LambdaZAP-II; 17-18 wk gestation, female; oligo-dT + random primed cDNA synthesis; lambdaZAP-II vector, 1.0kb average inser size."
BASE COUNT 137 a 58 c 68 g 106 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 370;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
||||||||||||||||
Db 113 CAAGTCACAGAATTA 127

RESULT 26
AI694429
LOCUS AI694429 402 bp mRNA EST 02-JUN-1999
DEFINITION wd83f06.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2338211 3', mRNA sequence.
ACCESSION AI694429
VERSION AI694429.1 GI:4971769
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.
FEATURES Location/Qualifiers
source 1. .402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2338211"
/clone_lib="NCI_CGAP_Lu24"
/tissue_type="carcinoid"
/lab_host="DH10B"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1522439). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 160 a 60 c 69 g 113 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
||||||||||||||||
Db 342 CAAGTCACAGAATTA 356

RESULT 27
AI203847
LOCUS AI203847 404 bp mRNA EST 27-JAN-1999
DEFINITION qf75h03.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1755893 3', mRNA sequence.
ACCESSION AI203847
VERSION AI203847.1 GI:3756453
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo

```
            , Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 563   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 394.
FEATURES             Location/Qualifiers
     source          1. .404
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1755893"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      161 a     61 c     69 g    113 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 404;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     342 CAAGTCACAGAATTA 356


RESULT  28
BF726110
LOCUS       BF726110      405 bp    mRNA            EST       05-JAN-2001
DEFINITION  by01c01.x1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
            sapiens cDNA clone by01c01 3', mRNA sequence.
ACCESSION   BF726110
VERSION     BF726110.1  GI:12042021
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
  TITLE     NEIBANK: EST analysis and bioinformatics for ocular genomics
  JOURNAL   Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 01  row: c  column: 01
            Seq primer: -21M13 forward primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .405
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="by01c01"
                     /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
                     BY"
                     /tissue_type="Lens"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
                     from different adults (both approximately 40 years old)
                     together yielded 20ug of total RNA and 150ng mRNA for cDNA
                     library synthesis. A directionally cloned cDNA library in
                     the pCMVSPORT6 vector was constructed at Life Technologies
                     , essentially following the protocols of the SuperScript
                     Plasmid System full details of which are contained in the
                     manufacturer's Instruction manual
                     (http://www.lifetech.com/). First strand synthesis was
                     carried out using a Not I primer-adapter
                     [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. Not I/blunt
                     end inserts were cloned into the Not I/EcoR V sites in the
                     vector. EST analysis was performed on the unamplified
                     library at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT      140 a     58 c     47 g    160 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 405;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     249 CAAGTCACAGAATTA 263


RESULT  29
BG413230/c
LOCUS       BG413230      406 bp    mRNA            EST       13-MAR-2001
DEFINITION  EtESTed18g03.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5' similar to SW:FBN1_BOVIN P98133 FIBRILLIN 1
            PRECURSOR ;, mRNA sequence.
ACCESSION   BG413230
VERSION     BG413230.1  GI:13318783
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
```

EcoRI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>=10,from 1682 previous reads). Insert sizes range from 1.2-2.9Kb. The library may contain a small percentage of host or bacterial contaminents."

BASE COUNT 91 a 126 c 114 g 75 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gacattgtcgcaagt 15
||||||||||||||||
Db 406 GACATTGTCGCAAGT 392

RESULT 30
AI942464
LOCUS AI942464 416 bp mRNA EST 06-SEP-1999
DEFINITION wp80c05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2468072 3', mRNA sequence.
ACCESSION AI942464
VERSION AI942464.1 GI:5707120
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 416)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTGAP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 410.

FEATURES Location/Qualifiers
source 1. .416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2468072"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATAGGTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 164 a 63 c 73 g 116 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 416;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
||||||||||||||||
Db 331 CAAGTCACAGAATTA 345

RESULT 31
AA172188
LOCUS AA172188 420 bp mRNA EST 09-MAR-1998
DEFINITION zp29f06.s1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA clone IMAGE:610883 3', mRNA sequence.
ACCESSION AA172188
VERSION AA172188.1 GI:1751265
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 420)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 934 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 294.

FEATURES Location/Qualifiers
source 1. .420
/organism="Homo sapiens"
/db_xref="GDB:4626142"
/db_xref="taxon:9606"
/clone="IMAGE:610883"
/clone_lib="Stratagene neuroepithelium (#937231)"
/dev_stage="Ntera-2/RA neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. NT2 cells (Ntera-2/cl.D1) induced with Retinoic Acid for 24 hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT 164 a 63 c 72 g 118 t 3 others
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 420;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caagtcacagaatta 25
||||||||||||||||
Db 338 CAAGTCACAGAATTA 352

RESULT 32
AQ142435/c
LOCUS AQ142435 427 bp DNA GSS 24-SEP-1998
DEFINITION HS_3064_A1_A11_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3064 Col=21 Row=A, DNA sequence.

```
ACCESSION   AQ142435
VERSION     AQ142435.1  GI:3533088
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3064  row: A  column: 21
            Class: BAC ends
            High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3064 Col=21 Row=A"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      137 a     78 c    103 g    108 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     198 GCAAGTCACAGAATT 184


RESULT  33
AW870882/c
LOCUS       AW870882      429 bp    mRNA            EST       10-MAY-2001
DEFINITION  ra51h09.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
            cDNA 5' similar to WP:C26D10.4 CE03027 ;, mRNA sequence.
ACCESSION   AW870882
VERSION     AW870882.1  GI:8004935
KEYWORDS    EST.
SOURCE      southern root-knot nematode.
  ORGANISM  Meloidogyne incognita
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.
            , Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.
            , Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R.
            , Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
            ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.
            , Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Uma Rao and David Bird
            (david_bird@ncsu.edu) at North Carolina State University. DNA
            Sequencing by: Washington University Genome Sequencing Center St.
            Louis.
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 428.
FEATURES             Location/Qualifiers
     source          1. .429
                     /organism="Meloidogyne incognita"
                     /db_xref="taxon:6306"
                     /clone_lib="Bird-Rao Meloidogyne incognita J2"
                     /dev_stage="enriched for 2nd stage juveniles"
                     /lab_host="XLORL"
                     /note="Vector: ZAP express - pBKCMV (Stratagene); Site_1:
                     EcoRI; Site_2: XhoI; Oligo (dT) primed library.  cDNA was
                     contructed and cloned  unidirectionally into the vector
                     within the 5' EcoRI and  3' XhoI sites.  This library was
                     constructed by Dr. Uma Rao  and Dr. David Bird at North
                     Carolina state University."
BASE COUNT      127 a     66 c     38 g     198 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 429;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     385 CAAGTCACAGAATTA 371


RESULT  34
AQ086771
LOCUS       AQ086771      454 bp    DNA             GSS       26-AUG-1998
DEFINITION  HS_2175_B2_A06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2175 Col=12 Row=B, DNA sequence.
ACCESSION   AQ086771
VERSION     AQ086771.1  GI:3455988
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2175  row: B  column: 12
            Class: BAC ends
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1. .454
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2175 Col=12 Row=B"
```

```
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       188 a     70 c     76 g    117 t      3 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 454;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     156 CAAGTCACAGAATTA 170


RESULT  35
AA973493
LOCUS       AA973493      455 bp    mRNA            EST       17-JUN-1998
DEFINITION  oo45e05.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1569152 3',
            mRNA sequence.
ACCESSION   AA973493
VERSION     AA973493.1  GI:3148673
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 553   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 440.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1569152"
                     /clone_lib="NCI_CGAP_Lu5"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     neuroendocrine lung carcinoid, and was then primed with a
                     Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                     to Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector.  Library is normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       175 a     64 c     84 g    132 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     340 CAAGTCACAGAATTA 354


RESULT  36
AI756736/c
LOCUS       AI756736      456 bp    mRNA            EST       18-JAN-2000
DEFINITION  EtESTea24h07.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA
            5' similar to SW:FBN1_BOVIN P98133 FIBRILLIN 1 PRECURSOR ;, mRNA
            sequence.
ACCESSION   AI756736
VERSION     AI756736.1  GI:5150459
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria S5-2 Sporozoite stage"
                     /dev_stage="Sporozoite"
                     /lab_host="SOLR E. coli"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E. tenella grown in chickens.  cDNA
                     was synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to  the cDNA and products were
                     size-selected on Sephacryl S500.  cDNAs were digested with
                     EcoRI/XhoI and cloned into lambda Zap II (Stratagene).
                     Clones were converted to phagemids by mass excision  using
                     ExAssist helper phage and SOLR cells (Stratagene ).
                     Insert  sizes range from 1.2-2.9 kb."
BASE COUNT       100 a     137 c     132 g     87 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagt 15
           |||||||||||||||
Db     435 GACATTGTCGCAAGT 421


RESULT  37
AI361386
LOCUS       AI361386      457 bp    mRNA            EST       15-FEB-1999
DEFINITION  qy47f01.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2015161 3',
            mRNA sequence.
ACCESSION   AI361386
```

```
VERSION     AI361386.1  GI:4113007
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1228   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 434.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2015161"
                     /clone_lib="NCI_CGAP_Brn23"
                     /tissue_type="glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATATCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      153 a      73 c      52 g     179 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 457;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     153 CAAGTCACAGAATTA 167


RESULT  38
BG562271/c
LOCUS       BG562271      464 bp    mRNA            EST       10-APR-2001
DEFINITION  EtESTee05a02.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5' similar to SW:FBN1_BOVIN P98133 FIBRILLIN 1
            PRECURSOR ;, mRNA sequence.
ACCESSION   BG562271
VERSION     BG562271.1  GI:13591269
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT      101 a     141 c     137 g      85 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagt 15
           |||||||||||||||
Db     421 GACATTGTCGCAAGT 407


RESULT  39
AI635259
LOCUS       AI635259      471 bp    mRNA            EST       16-DEC-1999
DEFINITION  tz79h07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2294845 3',
            mRNA sequence.
ACCESSION   AI635259
VERSION     AI635259.1  GI:4686589
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```

```
          www-bio.llnl.gov/bbrp/image/image.html
          Insert Length: 1533   Std Error: 0.00
          Seq primer: -40UP from Gibco
          High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2294845"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT      171 a     69 c     53 g    178 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 471;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     235 CAAGTCACAGAATTA 249


RESULT  40
BG020076
LOCUS       BG020076      485 bp    mRNA            EST       24-JAN-2001
DEFINITION  dc47c07.x1 NICHD XGC Emb3 Xenopus laevis cDNA clone IMAGE:3400308
            3' similar to TR:Q9VCA6 Q9VCA6 CG6313 PROTEIN. ;, mRNA sequence.
ACCESSION   BG020076
VERSION     BG020076.1  GI:12476155
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 409.
FEATURES             Location/Qualifiers
     source          1. .485
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:3400308"
                     /clone_lib="NICHD XGC Emb3"
                     /tissue_type="embryo (stages 24-25)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 1.7 kb. Constructed by Life Technologies. Note: This
                     is a Xenopus Gene Collection (XGC) library."
BASE COUNT      117 a    102 c     91 g    165 t     10 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 485;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     437 GCAAGTCACAGAATT 451


RESULT  41
BG516037/c
LOCUS       BG516037      487 bp    mRNA            EST       30-MAR-2001
DEFINITION  EtESTed48c02.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria
            tenella cDNA 5' similar to SW:FBN1_BOVIN P98133 FIBRILLIN 1
            PRECURSOR ;, mRNA sequence.
ACCESSION   BG516037
VERSION     BG516037.1  GI:13489363
KEYWORDS    EST.
SOURCE      Eimeria tenella.
  ORGANISM  Eimeria tenella
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
            Eimeria.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T.,
            Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen
            ,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey
            ,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson
            ,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU-Merck Eimeria tenella project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Sibley, Ph.D.
            WashU-Merck Eimeria tenella project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Contact David Sibley (toxoest@borcim.wustl.edu) for further
            information relating to organism, libraries, or clone availability.
            Seq primer: -40RP from Gibco
            High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Eimeria tenella"
                     /strain="LS18"
                     /db_xref="taxon:5802"
                     /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected"
                     /dev_stage="Sporozoite stage"
                     /lab_host="SOLR"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Sporozoites were obtained from in vitro sporulated and
                     excysted oocysts of E.tenella grown in chickens. cDNA was
                     synthesized from poly mRNA using an oligo-dT primer
                     containing a XhoI site. Following second strand synthesis,
                     EcoRI adapters were ligated to the cDNA and products were
                     size-selected on sephacryl S500. The cDNA were ligated to
                     EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were
                     converted to phagemids by mass excision using ExAssist
                     helper phage and E.coli SOLR cell (Stratagene). Clones
                     were selected by negative hybridization against a pool of
                     overrepresented ESTs (N>=10,from 1682 previous reads).
                     Insert sizes range from 1.2-2.9Kb. The library may contain
                     a small percentage of host or bacterial contaminents."
BASE COUNT      105 a    147 c    142 g     93 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 487;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gacattgtcgcaagt 15
           |||||||||||||||
```

```
Db    451 GACATTGTCGCAAGT 437


RESULT  42
BF775935/c
LOCUS       BF775935      489 bp    mRNA            EST       25-APR-2001
DEFINITION  286217 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF775935
VERSION     BF775935.1  GI:12123835
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 88  row: K  column: 20
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT      147 a     85 c     95 g    162 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 489;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           |||||||||||||||
Db     153 CAAGTCACAGAATTA 139


RESULT  43
AZ398330
LOCUS       AZ398330      498 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0163N15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0163N15 R, DNA sequence.
ACCESSION   AZ398330
VERSION     AZ398330.1  GI:10513498
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0163  row: N  column: 15
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 498.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0163N15"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      187 a     91 c    117 g    103 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 498;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gcaagtcacagaatt 24
           |||||||||||||||
Db     101 GCAAGTCACAGAATT 115


RESULT  44
AA846453
LOCUS       AA846453      502 bp    mRNA            EST       31-DEC-1998
DEFINITION  aj56b03.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394285
            3', mRNA sequence.
ACCESSION   AA846453
VERSION     AA846453.1  GI:2932593
KEYWORDS    EST.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 925    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 448.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1394285"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      196 a     67 c     97 g    142 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 502;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||| 
Db     342 CAAGTCACAGAATTA 356


RESULT  45
AI672463
LOCUS       AI672463       502 bp    mRNA            EST       16-DEC-1999
DEFINITION  wa03e05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2297024 3',
            mRNA sequence.
ACCESSION   AI672463
VERSION     AI672463.1  GI:4852194
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 535    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 458.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2297024"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo.  "
BASE COUNT      198 a     67 c     96 g    141 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 502;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 caagtcacagaatta 25
           ||||||||||||||| 
Db     339 CAAGTCACAGAATTA 353


Search completed: December 26, 2001, 14:09:53
Job time: 7040 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:53:44 ; Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 226708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		25	100.0	1041	6	AR029499	AR029499 Sequence
2		25	100.0	1041	6	AR034916	AR034916 Sequence
3		25	100.0	1084	6	A11530	A11530 BioB gene o
4		25	100.0	1121	6	E00893	E00893 Genomic DNA
5		25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
6		25	100.0	5872	6	A38246	A38246 Sequence 1
7		25	100.0	5872	6	A38251	A38251 Sequence 6
8		25	100.0	5872	6	A93674	A93674 Sequence 1
9		25	100.0	5872	6	A93679	A93679 Sequence 6
10		25	100.0	5872	6	AR101809	AR101809 Sequence
11		25	100.0	5872	6	AR101810	AR101810 Sequence
12		25	100.0	11022	1	AE000180	AE000180 Escherich
13		21	84.0	7215	1	SMABIO	D17468 S. marcesce
14		21	84.0	13501	1	AE005258	AE005258 Escherich
15		21	84.0	297816	1	AP002553	AP002553 Escherich
16	c	19	76.0	213732	1	AE001862	AE001862 Deinococc
17		17	68.0	2081	10	HAMHPP	M96650 Mesocricetu
18	c	17	68.0	7318	10	AB026497	AB026497 Mus muscu
19	c	17	68.0	69900	2	AC021767	AC021767 Mus muscu
20		17	68.0	151340	2	AC069132	AC069132 Homo sapi
21	c	17	68.0	161742	2	AC068100	AC068100 Homo sapi
22	c	17	68.0	168360	2	AC041048	AC041048 Homo sapi
23	c	17	68.0	176379	2	AC060833	AC060833 Homo sapi
24	c	17	68.0	220166	2	AC022781	AC022781 Mus muscu
25		17	68.0	222212	2	AL591065	AL591065 Mus muscu
26		16	64.0	8326	5	AF237379	AF237379 Xenopus l
27		16	64.0	10536	1	AE004050	AE004050 Xylella f
28		16	64.0	12242	1	AE005241	AE005241 Escherich
29		16	64.0	38065	1	MLCB1450	AL035159 Mycobacte
30		16	64.0	39520	2	AC091120	AC091120 Leishmani
31		16	64.0	71230	9	AC004508	AC004508 Homo sapi
32	c	16	64.0	85633	9	AC002088	AC002088 Human BAC
33	c	16	64.0	109973	2	AC007555	AC007555 Homo sapi
34	c	16	64.0	119749	2	AC092845	AC092845 Homo sapi
35	c	16	64.0	127070	9	AC073052	AC073052 Homo sapi
36	c	16	64.0	155628	2	AC027810	AC027810 Homo sapi
37		16	64.0	156119	2	AC073247	AC073247 Homo sapi
38		16	64.0	162043	2	AC040930	AC040930 Homo sapi
39	c	16	64.0	175459	2	AC068538	AC068538 Homo sapi
40	c	16	64.0	175588	9	AL139326	AL139326 Human DNA
41	c	16	64.0	182396	2	AC092679	AC092679 Homo sapi
42	c	16	64.0	183700	9	AC005412	AC005412 Homo sapi
43	c	16	64.0	185492	33	AC025339	Ac025339 Homo sapi
44		16	64.0	202828	33	AC064835	Ac064835 Homo sapi
45	c	16	64.0	245499	2	AC016186	AC016186 Homo sapi
46		16	64.0	262278	1	AP002552	AP002552 Escherich
47	c	16	64.0	294800	1	SME591789	AL591789 Sinorhizo
48		16	64.0	313064	9	HSMX1B	AL442167 Homo sapi
49		16	64.0	340000	9	HS21C085	AL163285 Homo sapi
50		15	60.0	128	1	ECOBIOB	M27731 Escherichia
51		15	60.0	132	11	G47166	G47166 Z15776_1 Ze
52	c	15	60.0	188	9	HSIGCOS1	Z17387 H.sapiens g
53	c	15	60.0	247	10	RNO277664	AJ277664 Rattus no
54	c	15	60.0	247	10	RNO277667	AJ277667 Rattus no
55	c	15	60.0	253	10	RNTCRAV18	Y09186 R.norvegicu
56	c	15	60.0	273	10	RNTCRV83	X80510 R.norvegicu
57	c	15	60.0	630	9	HUMIGH210P	M99647 Human immun
58		15	60.0	636	9	AF072252	AF072252 Homo sapi
59		15	60.0	885	2	AC075977	AC075977 Giardia i
60	c	15	60.0	996	2	AC076395	AC076395 Giardia i
61	c	15	60.0	1065	2	AC072114	AC072114 Giardia i
62	c	15	60.0	1122	2	AC071790	AC071790 Giardia i
63	c	15	60.0	1161	2	AC077117	AC077117 Giardia i
64		15	60.0	1552	5	XLHNF4MR	Z37526 X.laevis HN
65	c	15	60.0	1727	9	AB056427	AB056427 Macaca fa
66		15	60.0	1938	3	AF299082	AF299082 Giardia i
67		15	60.0	2088	10	AF238293	AF238293 Mus muscu
68		15	60.0	2931	5	DRAJ5029	AJ005029 Danio rer
69		15	60.0	3829	9	HSMETHYL7	AF120994 Homo sapi
70	c	15	60.0	4380	9	HSA317970	AJ317970 Homo sapi

c	71	15	60.0	4674	9	AF034373	AF034373 Homo sapi
	72	15	60.0	5526	1	AF250776	AF250776 Unculture
	73	15	60.0	8482	9	AF281036	AF281036 Homo sapi
	74	15	60.0	12891	1	AE004192	AE004192 Vibrio ch
c	75	15	60.0	15143	9	HSAJ4862	AJ004862 Homo sapi
	76	15	60.0	20780	2	AC014847	AC014847 Drosophil
c	77	15	60.0	22773	9	AF107890	AF107890 Homo sapi
	78	15	60.0	23185	2	AC014149	AC014149 Drosophil
	79	15	60.0	32886	9	HSU201H11	Z69721 Human DNA s
c	80	15	60.0	39017	9	HSJ850N15	AL121831 Human DNA
	81	15	60.0	44706	2	AC022328	AC022328 Mus muscu
c	82	15	60.0	50149	2	AC011955	AC011955 Homo sapi
c	83	15	60.0	61010	9	AL138961	AL138961 Human DNA
c	84	15	60.0	61170	2	AC068163	AC068163 Homo sapi
c	85	15	60.0	68334	2	AC010563	AC010563 Drosophil
	86	15	60.0	73167	9	AC010316	AC010316 Homo sapi
	87	15	60.0	79288	3	AE003572	AE003572 Drosophil
c	88	15	60.0	79938	2	AC025399	AC025399 Homo sapi
	89	15	60.0	87417	2	AL389896	AL389896 Homo sapi
	90	15	60.0	92588	9	HSIMMDL	X97051 DNA sequenc
c	91	15	60.0	102326	9	AC087407	AC087407 Homo sapi
	92	15	60.0	106482	10	AC007585	AC007585 Mus muscu
	93	15	60.0	110000	2	AC087782_2	Continuation (3 of
c	94	15	60.0	110000	2	AL591074_4	Continuation (5 of
c	95	15	60.0	110000	2	LMFLCHR31_11	Continuation (12 o
c	96	15	60.0	118561	2	AC009217	AC009217 Drosophil
	97	15	60.0	122914	10	AC069451	AC069451 Mus muscu
c	98	15	60.0	136804	9	AP000469	AP000469 Homo sapi
c	99	15	60.0	137889	9	AC073269	AC073269 Homo sapi
	100	15	60.0	141704	9	AC011754	AC011754 Homo sapi

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT       262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.0027;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    111 TCGTCAGGTGCAGGTCAGCACGTTG 135


RESULT   2
AR034916
LOCUS       AR034916      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT       262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 0.0027;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    111 TCGTCAGGTGCAGGTCAGCACGTTG 135


RESULT   3
A11530
LOCUS       A11530        1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT       271 a    286 c    318 g    209 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 0.0027;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    134 TCGTCAGGTGCAGGTCAGCACGTTG 158


RESULT   4
E00893
LOCUS       E00893        1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
```

```
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1121;
  Best Local Similarity  100.0%;  Pred. No. 0.0027;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    152 TCGTCAGGTGCAGGTCAGCACGTTG 176


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
```

```
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match             100.0%;  Score 25;  DB 1;  Length 5793;
  Best Local Similarity   100.0%;  Pred. No. 0.002;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db    2122 TCGTCAGGTGCAGGTCAGCACGTTG 2146


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
```

```
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.002;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.002;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
```

```
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAI
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN
```

```
Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
Best Local Similarity   100.0%;  Pred. No. 0.002;
Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251
```

```
RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA              PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN
```

```
Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
Best Local Similarity   100.0%;  Pred. No. 0.002;
```

```
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT  10
AR101809
LOCUS       AR101809     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.002;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT  11
AR101810
LOCUS       AR101810     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.002;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT  12
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
```

```
CDS             complement(147. .1430)
                /gene="ybhC"
                /function="orf; Not classified"
                /note="f427; 98 pct identical to fragment YBHC_ECOLI
                SW:P46130 (300 aa) but contains 127 additional C-terminal
                residues"
                /codon_start=1
                /transl_table=11
                /product="putative pectinesterase"
                /protein_id="AAC73859.1"
                /db_xref="GI:1786989"
                /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHETAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
```

```
  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 0.0018;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db   3603 TCGTCAGGTGCAGGTCAGCACGTTG 3627


RESULT  13
SMABIO
LOCUS       SMABIO       7215 bp    DNA             BCT       04-FEB-1999
DEFINITION  S. marcescens biotin operon, complete sequence.
ACCESSION   D17468
VERSION     D17468.1  GI:402530
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-aminopelargonic acid synthetase; DAPA aminotransferase;
            DTB synthetase; KAPA synthetase; biotin operon; biotin synthetase;
            dethiobiotin synthetase.
SOURCE      Serratia marcescens (strain:Sr41) DNA.
  ORGANISM  Serratia marcescens
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
```

```
            Serratia.
REFERENCE   1  (bases 1 to 7215)
  AUTHORS   Sakurai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki
            Sakurai, Tanabe Seiyaku Co., Ltd., Res Lab of Applied Biochemistry;
            2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan
            (E-mail:GFC01101@niftyserve.or.jp, nsakurai@ddbj.nig.ac.jp,
            Tel:048-433-2545, Fax:048-433-2540)
REFERENCE   2  (bases 1 to 7215)
  AUTHORS   Sakurai,N., Imai,Y., Akatsuka,H., Kawai,E., Komatsubara,S. and
            Tosa,T.
  TITLE     Complete nucleotide sequence of biotin operon of Serratia
            marcescens
  JOURNAL   Unpublished (1993)
COMMENT     Submitted (25-Aug-1993) to DDBJ by:
            Naoki Sakurai
            Res. Lab. of Applied Biochemistry
            Tanabe Seiyaku Co., Ltd.
            2-50 Kawagishi-2-chome
            Toda, Saitama 335
            Japan
            Phone: 048-433-2545
            Email: nsakurai@ddbj.nig.ac.jp
            Fax:   048-433-2540.
FEATURES             Location/Qualifiers
     source          1. .7215
                     /organism="Serratia marcescens"
                     /strain="Sr41"
                     /db_xref="taxon:615"
     gene            complement(1353. .2630)
                     /gene="bioA"
     CDS             complement(1353. .2630)
                     /gene="bioA"
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase
                     (DAPA aminotransferase)"
                     /protein_id="BAA04284.1"
                     /db_xref="GI:402531"
                     /translation="MSVTASDLAFDQRHIWHPYTSMSRPLPCYPIESASGVELQLADG
                     RSLVDGMSSWWAAIHGYNHPRLNQAASQQLEKMSHVMFGGITHPAAISLCRRLVEMTP
                     EALQCVFLADSGSVAVEVSLKMALQYWQARGERRQRILTLRHGYHGDTFGAMSVCDPD
                     NSMHSLYQGYLAPHLFATAPQCRFDEEWREEDIAPFAALLEQHAGEVAAVILEPVVQG
                     AGGMRIYHPTYLKRVRDVVRVAIKLLLIADEIATGFGRTGKLFACEHAQVVPDILCLG
                     KALTGGYMTLSATLTTRHVAETISNGAAGCFMHGPTFMGNPLACAVADASLALLAENR
                     WQAQVSAIETQLKRELLPLGRCRRWPHVRVLGAIGVVEMREPVDVAELQRGFVERGVW
                     IRPFGKLIYLMPPYIIEAEQLSRLTAAVADAAR"
     gene            2716. .3756
                     /gene="bioB"
     CDS             2716. .3756
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="BAA04285.1"
                     /db_xref="GI:402532"
                     /translation="MMADRIHWTVGQAQALFDKPLLELLFEAQTVHRQHFDPRQVQVS
                     TLLSIKTGACPEDCKYCPQSPRYKTGLESERLMQVEQVLESARKAKANGSTRFCMGAA
                     WKNPHERDMPYLQQMVQGVKAMGMETCMTLGTLDGTQAERLAEAGLDYYNHNLDTSPE
                     FYGSIITTRSYQERLDTLDKVRDAGIKVCSGGIVGLGETVRDRAGLLVQLANLPKPPE
                     SVPINMLVKVKGTPLADNDDVDPFDFIRTIAVARIMMPSSYVRLSAGREQMNEQTQAM
                     CFMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTATEHGDNQQQQVLAKQLL
                     NADTAEFYNAAP"
     gene            3756. .4904
                     /gene="bioF"
     CDS             3756. .4904
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-aminopelargonic acid synthetase (KAPA
                     synthetase)"
                     /protein_id="BAA04286.1"
                     /db_xref="GI:402533"
                     /translation="MSWQQRIEQALAERRLNAAYRRRQTTEGGNGRQIRLGDRLYLNF
                     SGNDYLGLSQDARVIAAWQQGARYGVGNGGSGHVTGFSAAHQALEEQLAAWLGYPRAL
                     LFISYAANQAVLAALMQKGDRILADRLSHASLLEAAGAVAGRRAPVPAQSTAGLARIC
                     LAKPCDGQRLAVTEGLFSMDGDGAPLAELHRLTRAAGAWLMVDDAHGIGVRGEQGRGS
                     CWQQGVRPELLVATFGKAFGVSGAAVLCDEATAEYLLQFARHLIYSTAMPPAQACALQ
                     AALARIREGDDLRARLQDNIRRFRQGAAPLALTLTDSDTAIQPLLVGDNQRALDLATR
                     LRECGLWVSAIRPPTVPPGGARLRLLLTAAHQSQDIDRLLEVLNDVSQ"
     gene            4888. .5655
                     /gene="bioC"
     CDS             4888. .5655
                     /gene="bioC"
                     /function="synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="the product of bioC"
                     /protein_id="BAA04287.1"
                     /db_xref="GI:402534"
                     /translation="MTSANDTVNKQAVASAFSRAAGSYDAAAALQRDVGERLLGMGSS
                     HPGEQLLDAGCGTGYFSRMWRERGKRVTALDLAPGMLDVARQRQAAHHYLLGDIEQVP
                     LPDAAMDICFSSLVVQWCSDLPAALAELYRVTRPGGVILFSTLAAGSLQELGDAWQQV
                     DGERHVNAFLPLTQIRTACAAYRHELVTELRTLNYPDVMTLMRSLKGIGATHLHQGRE
                     GGLMSRGRLAALQAAYPCRQGQFPLSYHLAYGVIYRE"
     gene            5648. .6331
                     /gene="bioD"
     CDS             5648. .6331
                     /gene="bioD"
                     /function="conversion from 7,8-diaminopelargonic acid into
                     dethiobiotin"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase(DTB synthetase)"
                     /protein_id="BAA04288.1"
                     /db_xref="GI:912453"
                     /translation="MSKRWFVTGTDTEVGKSVASSALLQAANRAGYPSAGYKPVASGS
                     EMTAEGLRNGDALALQANSGVALDYDEVNPYVFAEPTSPHIVSADEGRPIDAARLSDG
                     LRRLEQRADWVLVEGAGGWFTPLSAEYTFADWVRQEQLPVILVVGIKLGCINHAVLTA
                     QAVQQAGLTLAGWIANDVGGAPGRRHQEYLATLRRMLPRRCWAKSRTCRRPERAPLGQ
                     YLDISLLAQ"
BASE COUNT     1347 a   2225 c   2367 g   1276 t
ORIGIN


  Query Match             84.0%;  Score 21;  DB 1;  Length 7215;
  Best Local Similarity   100.0%;  Pred. No. 0.25;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcagcacg 22
           |||||||||||||||||||||
Db    2830 CGTCAGGTGCAGGTCAGCACG 2850


RESULT  14
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
```

```
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSCDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
```

```
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match             84.0%;  Score 21;  DB 1;  Length 13501;
  Best Local Similarity   100.0%;  Pred. No. 0.22;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcac 21
          |||||||||||||||||||||
Db   9689 TCGTCAGGTGCAGGTCAGCAC 9709


RESULT  15
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase alpha subunit"
                     /protein_id="BAB34177.1"
                     /db_xref="GI:13360213"
                     /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                     GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
```

```
                LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                LADIGEALKTVLK"
     gene       2781. .3686
                /gene="ECs0755"
     CDS        2781. .3686
                /gene="ECs0755"
                /note="probable transcriptional regulator,similar to
                transcriptional regulators e.g. glycine cleavage system
                transcription activator (gcv operon activator) -
                Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                identity 31 in 300 aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transcription regulator"
                /protein_id="BAB34178.1"
                /db_xref="GI:13360214"
                /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                EKSDSPHLQQFREWLLAKSVVPQEM"
     gene       complement(3720. .4322)
                /gene="ECs0756"
     CDS        complement(3720. .4322)
                /gene="ECs0756"
                /note="probable cob(I)alamin adenosyltransferase,similar
                to cob(I)alamin adenosyltransferases (corrinoid
                adenosyltransferases) e.g. [Escherichia coli]
                gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative cob(I)alamin adenosyltransferase"
                /protein_id="BAB34179.1"
                /db_xref="GI:13360215"
                /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
     gene       complement(4332. .5984)
                /gene="ECs0757"
     CDS        complement(4332. .5984)
                /gene="ECs0757"
                /note="probable fumarate hydratase,similar to fumarate
                hydratases e.g. fumarate hydratase class I, aerobic
                (fumarase) - Escherichia coli
                gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative fumarate hydratase"
                /protein_id="BAB34180.1"
                /db_xref="GI:13360216"
                /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
     gene       complement(6092. .7372)
                /gene="ECs0758"
     CDS        complement(6092. .7372)
                /gene="ECs0758"
                /note="probable transport protein,similar to
                glutamate/aspartate transport proteins (proton glutamate
                symport proteins) e.g. [Bacillus stearothermophilus]
                gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                aa,also similar to C4-dicarboxylate transpor"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transport protein"
                /protein_id="BAB34181.1"
                /db_xref="GI:13360217"
                /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
     gene       complement(7533. .7850)
                /gene="ECs0759"
     CDS        complement(7533. .7850)
                /gene="ECs0759"
                /note="unknown"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="hypothetical protein"
                /protein_id="BAB34182.1"

  Query Match             84.0%;  Score 21;  DB 1;  Length 297816;
  Best Local Similarity   100.0%;  Pred. No. 0.13;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcac 21
           |||||||||||||||||||||
Db   92421 TCGTCAGGTGCAGGTCAGCAC 92441


RESULT  16
AE001862/c
LOCUS       AE001862   213732 bp    DNA             BCT       22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 1 of 2 of the complete
            chromosome 2.
ACCESSION   AE001862 AE001825
VERSION     AE001862.1  GI:6460468
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
REFERENCE   1  (bases 1 to 213732)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et,al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 213732)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .213732
```

```
                /organism="Deinococcus radiodurans"
                /strain="R1"
                /db_xref="taxon:1299"
                /chromosome="2"
     gene       653. .1435
                /gene="DRA0001"
     CDS        653. .1435
                /gene="DRA0001"
                /note="similar to GB:D26185 SP:P37522 PID:467381
                PID:580906 GB:AL009126 percent identity: 54.96; identified
                by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="chromosome partitioning ATPase, putative, ParA
                family"
                /protein_id="AAF12301.1"
                /db_xref="GI:6460595"
                /translation="MVSAVKTLTVFNHAGGAGKTSLTLNVGYELARGGLRVLLLDLDP
                QANLTGWLGISGVTREMTVYPVAVDGQPLPSPVKAFGLDVIPAHVSLAVAEGQMMGRV
                GAQGRLRRALAEVSGDYDVALIDSPPSLGQLAILAALAADQMIVPVPTRQKGLDALPG
                LQGALTEYREVRPDLTVALYVPTFYDARRRHDQEVLADLKAHLSPLARPVPQREAVWL
                DSTAQGAPVSEYAPGTPVHADVQRLTADIAAAIGVAYPGENA"
     gene       1432. .2313
                /gene="DRA0002"
     CDS        1432. .2313
                /gene="DRA0002"
                /note="similar to GB:L34077 GB:X85964 PID:407375
                PID:757761 PID:1163135 percent identity: 57.34; identified
                by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="chromosome partitioning protein, ParB family"
                /protein_id="AAF12300.1"
                /db_xref="GI:6460594"
                /translation="MTRRRPERRRDLLGLLGETPVDLSQANDIRALPVNELKVGSTQP
                RRSFDLERLSELAESIRAHGVLQPLLVRSVDGQYEIVAGERRWRAAQLAGLAEVPVVV
                RQLSNEQARAAALIENLQRDNLNVIDEVDGKLELIALTLGLEREEARKRLMQLLRAVP
                GDEHEQLDQVFRSMGETWRTFAKNKLRILNWPQPVLEALRAGLPLTLGSVVASAPPER
                QAELLKLAQNGASRSQLLQALQTPSQTSAVTPEHFAKVLSSKRFLSGLDTPTREALDR
                WLARMPERVRQAIDEQS"
     gene       complement(2418. .3881)
                /gene="DRA0003"
     CDS        complement(2418. .3881)
                /gene="DRA0003"
                /note="similar to GB:M88334 SP:P25526 PID:147901 GB:U00096
                PID:1789015 percent identity: 69.60; identified by
                sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="succinate-semialdehyde dehydrogenase"
                /protein_id="AAF12294.1"
                /db_xref="GI:6460588"
                /translation="MTTTDLRTTYSSVTRSQAYFDGEWRNAPRNFEVRHPGNGEVIGE
                VADCTPTDARQAIDAAEVALREWRQVNPYERGKILRRWHDLMFEHKEELAQLMTLEMG
                KPISETRGEVHYAASFIEWCAEEAGRIAGERINLRFPHKRGLTISEPVGIVYAVTPWN
                FPAGMITRKAAPALAAGCVMILKPAELSPMTALYLTELWLKAGGPANTFQVLPTNDAS
                ALTQPFMNDSRVRKLTFTGSTEVGRLLYQQAAGTIKRVSLELGGHAPFLVFDDADLER
                AASEVVASKFRNSGQTCVCTNRVYVQRGVAEEFIRLLTEKTAALQLGDPFDEATQVGP
                VVEQAGLDKVQRQVQDALTKGAQATTGGQVSSGLFFQPTVLVDVAPDSLILREETFGP
                VAPVTIFDTEEEGLRLANDSEYGLAAYAYTRDLGRAFRIAEGLEYGIVGINDGLPSSA
                APHVPFGGMKNSGVGREGGHWGLEEYLETKFVSLGLS"
     gene       4024. .4668
                /gene="DRA0004"
     CDS        4024. .4668
                /gene="DRA0004"
                /note="similar to GP:2924378 percent identity: 55.29;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF12299.1"
                /db_xref="GI:6460593"
                /translation="MNSLSSLAAVLLMAGRSTRMGQPKALLPLGGMPLCRHAAQTLVE
                AGYGEVWAVLPPGETGDAVRAALQGLPLRWVTNPQPAQGLLSSFQAAGEALWAESAAP
                AVVFTLADMPLVRRDTHQALGTVFGAEQPAAIITRYGEVAAPPTLLRRDLLPRLHEFP
                PSDHGPRTLLRELGRELKKIDRPAAELLDIDTPEAWAQAQQLFAEVTPGAQDGP"
     gene       complement(4719. .5897)
                /gene="DRA0005"
     CDS        complement(4719. .5897)
                /gene="DRA0005"
                /note="similar to GB:L33464 SP:P47734 PID:496118 percent
                identity: 74.23; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="alcohol dehydrogenase, zinc-containing"
                /protein_id="AAF12293.1"
                /db_xref="GI:6460587"
                /translation="MKAIVWQGVNKVGVETVPDPTLLLPTDAIVKITSTAICGSDLHL
                LDGYIPSMETGDILGHEFMGEVVEVGSEVRKLKVGDRVVVPFNIACGVCDPCKRGFFS
                ACDNSNPNHRMAEALYGATSGGALFGYSHMYGGYAGGQAQYVRVPFADVGPFKIESGL
                KDEQVLFLTDIFPTGYQAAEQCNIVPGRDVVAVFGAGPVGQFAARSAQMLGAAHVIVI
                DRVPERLAMAEAAGCQTINYEKEDVLLSLREATGGRGPDHVIDAVGMEAHGHGPGSTM
                DKVQQNLKLTFDRITALRWAILSCAKGGTVSMPGVYGGLVDKMPIGAAFAKGLIFRMG
                QTHTHRHIPSLLGRIEAGQIDPSFVITHRASLDQAPELYKTFRDKQDCCIKVVLNPWA
                "
     gene       complement(5894. .6700)
                /gene="DRA0006"
     CDS        complement(5894. .6700)
                /gene="DRA0006"
                /note="similar to PID:1653488 percent identity: 48.46;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF12292.1"
                /db_xref="GI:6460586"
                /translation="MDRRSSPQLRTMTNWNREQPAPHAGATGPNSSPGATMQPLERSL
                TGAVGAALAAAGLRSRSGLEKVVLTGLGAGLIAVAARGQNPLATALKIEQNDEGETVV
                RDAVTIGKPAEQLYAVWRDLPGLPLLMTHLRSVEVLDDKRSRWTVEAPAPLGTVSWEA
                ELTADEPGKRIAWRSLPGARIENSGEVLFRPAPGARGTEVVVRLTYRPPGGSAGAVIA
                RMFNQEPSQQLRDDLMRFKREQELGHAPTAQGQSSGRAALTPPNKRGDLV"
     gene       6839. .7744
                /gene="DRA0007"
     CDS        6839. .7744
                /gene="DRA0007"
                /note="similar to GB:AL009126 percent identity: 62.98;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter, ATP-binding protein, putative"
                /protein_id="AAF12291.1"
                /db_xref="GI:6460585"
                /translation="MITTFEQVSKTYGHVTALSDFNLTLRTGELTALLGPNGAGKSTA
                IGLLLGLSAPSAGQVRVLGADPRRNDVRARIGAMPQESALPAGLTVREAVTLFASFYP
                APLGVDEALALADLGPVAGRRAAQLSGGQKRRLAFALAVVGDPELLLIDEPTTGMDAQ
                SRAAFWEAVTGLRARGRTILLTTHYLEEAERTADRVVVMNGGRILADDTPQGLRSGVG
                GARVSFVSDLVQAELERLPGVSAVQVDAAGRADLRTSVPEALLAALIGSGTTFSDLEV
                RRATLEEAYLQLTGPQDMTAVTRSA"
     gene       7861. .8652
                /gene="DRA0008"
     CDS        7861. .8652
                /gene="DRA0008"
                /note="similar to GB:AL009126 percent identity: 52.46;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF12290.1"
                /db_xref="GI:6460584"
                /translation="MTTLPNVPARRTSALPVLAQLTLGELRRLLRSPMFTVGAIGFPV
                MFFALFGLPAVQEYGATDPHVGPVILTQFAAYSLLSLALFSFGAAVATERSGGWLRLL
                RSSPLPVPLYFVSKTLAALAFGAVSLALLYAFAHFAGGVTLPLGLALLLAGKLLLGMI
                PLVALGLCIGFLASPQAAQILANILSVVMSFASGLFVPLDQLPGFVQQIAPLLPAYHV
                SQIATNTVSGQTASEPAHWLALAAFTLVFGTLAVWGLKRDESREG"
     gene       8809. .9819
```

```
                     /gene="DRA0009"
     CDS             8809. .9819
                     /gene="DRA0009"
                     /note="similar to GB:AL009126 percent identity: 55.80;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF12289.1"
                     /db_xref="GI:6460583"
                     /translation="MLFWGTLVAFVAVYVFTFGKRDPDDRWAYVGWLSALLMWALGDW
                     LYGPSALSFLVYGGSLIGWQRRQWVALGGAAINALLMVGRLLTLPERSQGDFIFLLFI
                     LASAYANHATYRGLRARRRLAQVQLEKEKLAADAERERIARDLHDLLGHTLSVIVLKS
                     ELAGKLAEKHPQRAAAEIREVERISREALSEVRAAVQGYRGSGLAAELARAKVALDAA
                     GIRLVITDTLPELPADLESTLAMVLREAVTNVVRHSGASELRLTLRGQANGFQLTLQD
                     DGRGGDAPEGTGLNSMRERLRAMGGTFERDGHTGTRLQAFVPLSAGTAGLKAPELLEK
                     TS"
     gene            9816. .10451
                     /gene="DRA0010"
     CDS             9816. .10451
                     /gene="DRA0010"
                     /note="similar to GB:AL009126 percent identity: 67.84;
```

Query Match 76.0%; Score 19; DB 1; Length 213732;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4 tcaggtgcaggtcagcacg 22
           ||||||||||||||||||||
Db   52231 TCAGGTGCAGGTCAGCACG 52213
```

```
RESULT  17
HAMHPP
LOCUS       HAMHPP       2081 bp    mRNA            ROD       27-APR-1993
DEFINITION  Mesocricetus auratus pregnancy protein (HPP), complete cds.
ACCESSION   M96650
VERSION     M96650.1  GI:191387
KEYWORDS    pregnancy protein; pregnancy protein 60kDa.
SOURCE      Mesocricetus auratus (strain Waterhouse) (library: lambda gt 11)
            female adult, 14 days pregnant liver (maternal) cDNA to mRNA.
  ORGANISM  Mesocricetus auratus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
            Mesocricetus.
REFERENCE   1  (bases 1 to 2081)
  AUTHORS   Park,C.G.
  TITLE     Cloning and sequencing of a new pregnancy marker protein from
            hamster liver
  JOURNAL   Thesis (1992) Biological Sciences, The Wichita State University
FEATURES             Location/Qualifiers
     source          1. .2081
                     /organism="Mesocricetus auratus"
                     /strain="Waterhouse"
                     /db_xref="taxon:10036"
                     /dev_stage="adult, 14 days pregnant"
                     /sex="female"
                     /tissue_type="liver (maternal)"
                     /tissue_lib="lambda gt 11"
     sig_peptide     68. .130
                     /gene="HPP"
     CDS             68. .1330
                     /gene="HPP"
                     /codon_start=1
                     /product="pregnancy protein 60 kDa"
                     /protein_id="AAA37078.1"
                     /db_xref="GI:191388"
                     /translation="MALRVAFILLVAAFCPAVLCQQDGTLGKETTVQEEQNNETKVDS
                     LTLASINTDFAFSLYKQLALKNPDKNVIFSPFSISFALAFLSLGASSNTLEEILEGLK
                     FNLTETPEADIHRGFGHLLRMLSQPGDQVQVSTSSAMFVEKRLQILAEFKKKARALYQ
                     AEASSTDFQKPHEAKKLINDYVSKKTHGKIKELMSDLDDQTSMVLVNYIYFKGKWKAP
                     FDPDFTYESDFYLDNKRSVKVPMMKHKDLTTPYFRDEELSSYVVEMRYRGNVSAMFIL
                     PDKGRMQQLESSLQPETLRRWKDSLRPRMIHELYVPKFSISTDYSMEGILSQLGIKEV
                     FSSQADLSGITGDKDLRVSKVVHKAVLDVGETGTEAAAATGLRMMGSALIMNPLRLRF
                     NKPFLMIIYNTNTQTPLFMAKVTNPKQN"
     gene            68. .1330
                     /gene="HPP"
     mat_peptide     131. .1327
                     /gene="HPP"
                     /product="pregnancy protein 60 kDa"
BASE COUNT      547 a    572 c    484 g    478 t
ORIGIN
```

Query Match 68.0%; Score 17; DB 10; Length 2081;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       5 caggtgcaggtcagcac 21
           |||||||||||||||||
Db     455 CAGGTGCAGGTCAGCAC 471
```

```
RESULT  18
AB026497/c
LOCUS       AB026497     7318 bp    mRNA            ROD       29-APR-2000
DEFINITION  Mus musculus MysPDZ mRNA for myosin containing PDZ domain, complete
            cds.
ACCESSION   AB026497
VERSION     AB026497.1  GI:7416031
KEYWORDS    MysPDZ; myosin containing PDZ domain.
SOURCE      Mus musculus (strain:C57BL6) adult tissue_lib:spleen cDNA to mRNA,
            clone_lib:Lambda ZipLox.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Furusawa,T., Ikawa,S., Yanai,N. and Obinata,M.
  TITLE     Isolation of a novel PDZ-containing myosin from hematopoietic
            supportive bone marrow stromal cell lines
  JOURNAL   Biochem. Biophys. Res. Commun. 270 (1), 67-75 (2000)
  MEDLINE   20200096
REFERENCE   2  (bases 1 to 7318)
  AUTHORS   Furusawa,T., Obinata,M. and Yanai,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-1999) to the DDBJ/EMBL/GenBank databases. Tadashi
            Furusawa, Institute of Development,Aging and Cancer,Tohoku
            University, Department of Cell Biology; 4-1,Seiryomaci,Aoba-Ku,
            Sendai, Miyagi 980-8575, Japan (E-mail:t-furu@idac.tohoku.ac.jp,
            Tel:81-22-717-8486, Fax:81-22-717-8488)
FEATURES             Location/Qualifiers
     source          1. .7318
                     /organism="Mus musculus"
                     /strain="C57BL6"
                     /db_xref="taxon:10090"
                     /clone_lib="Lambda ZipLox"
                     /dev_stage="adult"
                     /tissue_lib="spleen"
     gene            52. .6159
                     /gene="MysPDZ"
     CDS             52. .6159
                     /gene="MysPDZ"
                     /codon_start=1
                     /product="myosin containing PDZ domain"
                     /protein_id="BAA93660.1"
                     /db_xref="GI:7416032"
                     /translation="MFNLMKKDKDKDGGRKEKKEKKEKKERMSAAELRSLEEMSMRRG
                     FFNLNRSSKRESKTRLEISNPIPIKVASGSDLHLTDIDSDSNRGSIILDSGHLSTASS
                     SDDLKGEEGSFRGSVLQRAAKFGSLAKQNSQMIVKRFSFSQRSRDESASETSTPSEHS
                     AAPSPQVEVRTLEGQLMQHPGLGIPRPGPRSRVPELVTKRFPADLRLPALVPPPPPAL
                     RELELQRRPTGDFGFSLRRTTMLDRAPEGQAYRRVVHFAEPGAGTKDLALGLVPGDRL
                     VEINGQNVENKSRDEIVEMIRQSGDSVRLKVQPIPELSELSRSWLRTGEGHRREPADA
                     KTEEQIAAEEAWYETEKVWLVHRDGFSLASQLKSEELSLPEGKARVKLDHDGAILDVD
                     EDDIEKANAPSCDRLEDLASLVYLNESSVLHTLRQRYGASLLHTYAGPSLLVLSTRGA
                     PAVYSEKVMHMFKGCRREDMAPHIYAVAQTAYRAMLMSRQDQSIVLLGSSGSGKTTSF
                     QHLVQYLATIAGTSGTKVFSVEKWQALSTLLEAFGNSPTIMNGSATRFSQILSLDFDQ
```

```
                     AGQVASASIQTMLLEKLRVARRPASEATFNVFYYLLACGDATLRTELHLNHLAENNVF
                     GIVPLSKPEEKQKAAQQFSKLQAAMKVLAISPEEQKTCWLILASIYHLGAAGATKEAA
                     EAGRKQFARHEWAQKAAYLLGCSLEELSSAIFKHQLKGGTLQRSTSFRQGPEESGLGE
                     GTKLSALECLEGMASGLYSELFTLLISLVNRALKSSQHSLCSMMIVDTPGFQNPEWGG
                     SARGASFEELCHNYAQDRLQRLFHERTFLQELERYKEDNIELAFDDLEPVADDSVAAV
                     DQASHLVRSLAHADEARGLLWLLEEEALVPGATEDALLDRLFSYYGPQEGDKKGQSPL
                     LRSSKPRHFLLGHSHGTNWVEYNVAGWLNYTKQNPATQNAPRLLQDSQKKIISNLFLG
                     RAGSATVLSGSIAGLEGGSQLALRRATSMRKTFTTGMAAVKKKSLCIQIKLQVDALID
                     TIKRSKMHFVHCFLPVAEGWPGEPRSASSRRVSSSSELDLPPGDPCEAGLLQLDVSLL
                     RAQLRGSRLLDAMRMYRQGYPDHMVFSEFRRRFDVLAPHLTKKHGRNYIVVDEKRAVE
                     ELLESLDLEKSSCCLGLSRVFFRAGTLARLEEQRDEQTSRHLTLFQAACRGYLARQHF
                     KKRKIQDLAIRCVQKNIKKNKGVKDWPWWKLFTTVRPLIQVQLSEEQIRNKDEEIQQL
                     RSKLEKVEKERNELRLSSDRLETRISELTSEILTDERNTGESASQLLDAETAERLRTEK
                     EMKELQTQYDALKKQMEVMEMEVMEARLIRAAEINGEVDDDDAGGEWRLKYERAVREV
                     DFTKKRLQQELEDKMEVEQQSRRQLERRLGDLQADSDESQRALQQLKKKCQRLTAELQ
                     DTKLHLEGQQVRNHELEKKQRRFDSELSQAHEETQREKLQREKLQREKDMLLAEAFSL
                     KQQMEEKDLDIAGFTQKVVSLEAELQDISSQESKDEASLAKVKKQLRDLEAKVKDQEE
                     ELDEQAGSIQMLEQAKLRLEMEMERMRQTHSKEMESRDEEVEEARQSCQKKLKQMEVQ
                     LEEEYEDKQKALREKRELESKLSTLSDQVNQRDFESEKRLRKDLKRTKALLADAQIML
                     DHLKNNAPSKREIAQLKNQLEESEFTCAAAVKARKAMEVEMEDLHLQIDDIAKAKTAL
                     EEQLSRLQREKNEIQNRLEEDQEDMNELMKKHKAAVAQASRDMAQMNDLQAQIEESNK
                     EKQELQEKLQALQSQVEFLEQSMVDKSLVSRQEAKIRELETRLEFEKTQVKRLENLAS
                     RLKETMEKLTEERDQRAAAENREKEQNKRLQRQLRDTKEEMSELARKEAEASRKKHEL
                     EMDLESLEAANQSLQADLKLAFKRIGDLQAAIEDEMESDENEDLINSEGDSDVDSELE
                     DRVDGVKSWLSKNKGPSKAPSDDGSLKSSSPTSHWKPLAPDPSDDEHDPVDSISRPRF
                     SHSYLSDSDTEAKLTETSA"
     repeat_region   68. .179
                     /note="KE repeat"
     misc_feature    635. .981
                     /gene="MysPDZ"
                     /note="PDZ domain"
     misc_feature    1272. .3558
                     /gene="MysPDZ"
                     /note="myosin head"
     misc_binding    1544. .1565
                     /gene="MysPDZ"
                     /bound_moiety="ATP/GTP binding site"
     misc_feature    3602. .3665
                     /gene="MysPDZ"
                     /note="IQ domain"
     misc_feature    3777. .5865
                     /gene="MysPDZ"
                     /note="coiled-coil"
     polyA_signal    7279. .7284
     polyA_site      7303
BASE COUNT     1768 a   1954 c   2232 g   1364 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 7318;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcag 18
          ||||||||||||||||
Db    288 CGTCAGGTGCAGGTCAG 272


RESULT  19
AC021767/c
LOCUS       AC021767    69900 bp    DNA             HTG       13-JUL-2000
DEFINITION  Mus musculus chromosome 11 clone RP23-47D19 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC021767
VERSION     AC021767.2  GI:7144812
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 69900)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 11, clone RP23-47D19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 69900)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 3, 2000 this sequence version replaced gi:6721397.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3201
                Center clone name: 47_D_19
            --------------.
            * NOTE: This record contains 79 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *       1      787: contig of 787 bp in length
            *     788 887: gap of      100 bp
            *     888     1651: contig of 764 bp in length
            *    1652 1751: gap of      100 bp
            *    1752     2546: contig of 795 bp in length
            *    2547 2646: gap of      100 bp
            *    2647     3438: contig of 792 bp in length
            *    3439 3538: gap of      100 bp
            *    3539     4327: contig of 789 bp in length
            *    4328 4427: gap of      100 bp
            *    4428     5234: contig of 807 bp in length
            *    5235 5334: gap of      100 bp
            *    5335     6129: contig of 795 bp in length
            *    6130 6229: gap of      100 bp
            *    6230     7040: contig of 811 bp in length
            *    7041 7140: gap of      100 bp
            *    7141     7931: contig of 791 bp in length
            *    7932 8031: gap of      100 bp
            *    8032     8810: contig of 779 bp in length
            *    8811 8910: gap of      100 bp
            *    8911     9696: contig of 786 bp in length
            *    9697 9796: gap of      100 bp
            *    9797    10596: contig of 800 bp in length
            *   10597 10696: gap of      100 bp
            *   10697    11493: contig of 797 bp in length
            *   11494 11593: gap of      100 bp
            *   11594    12399: contig of 806 bp in length
```

```
*    12400 12499: gap of      100 bp
*    12500    13296: contig of 797 bp in length
*    13297 13396: gap of      100 bp
*    13397    14152: contig of 756 bp in length
*    14153 14252: gap of      100 bp
*    14253    15034: contig of 782 bp in length
*    15035 15134: gap of      100 bp
*    15135    15922: contig of 788 bp in length
*    15923 16022: gap of      100 bp
*    16023    16795: contig of 773 bp in length
*    16796 16895: gap of      100 bp
*    16896    17698: contig of 803 bp in length
*    17699 17798: gap of      100 bp
*    17799    18568: contig of 770 bp in length
*    18569 18668: gap of      100 bp
*    18669    19457: contig of 789 bp in length
*    19458 19557: gap of      100 bp
*    19558    20342: contig of 785 bp in length
*    20343 20442: gap of      100 bp
*    20443    21239: contig of 797 bp in length
*    21240 21339: gap of      100 bp
*    21340    22122: contig of 783 bp in length
*    22123 22222: gap of      100 bp
*    22223    22995: contig of 773 bp in length
*    22996 23095: gap of      100 bp
*    23096    23836: contig of 741 bp in length
*    23837 23936: gap of      100 bp
*    23937    24720: contig of 784 bp in length
*    24721 24820: gap of      100 bp
*    24821    25590: contig of 770 bp in length
*    25591 25690: gap of      100 bp
*    25691    26480: contig of 790 bp in length
*    26481 26580: gap of      100 bp
*    26581    27368: contig of 788 bp in length
*    27369 27468: gap of      100 bp
*    27469    28258: contig of 790 bp in length
*    28259 28358: gap of      100 bp
*    28359    29134: contig of 776 bp in length
*    29135 29234: gap of      100 bp
*    29235    30016: contig of 782 bp in length
*    30017 30116: gap of      100 bp
*    30117    30897: contig of 781 bp in length
*    30898 30997: gap of      100 bp
*    30998    31770: contig of 773 bp in length
*    31771 31870: gap of      100 bp
*    31871    32662: contig of 792 bp in length
*    32663 32762: gap of      100 bp
*    32763    33558: contig of 796 bp in length
*    33559 33658: gap of      100 bp
*    33659    34448: contig of 790 bp in length
*    34449 34548: gap of      100 bp
*    34549    35346: contig of 798 bp in length
*    35347 35446: gap of      100 bp
*    35447    36251: contig of 805 bp in length
*    36252 36351: gap of      100 bp
*    36352    37136: contig of 785 bp in length
*    37137 37236: gap of      100 bp
*    37237    38021: contig of 785 bp in length
*    38022 38121: gap of      100 bp
*    38122    38908: contig of 787 bp in length
*    38909 39008: gap of      100 bp
*    39009    39800: contig of 792 bp in length
*    39801 39900: gap of      100 bp
*    39901    40683: contig of 783 bp in length
*    40684 40783: gap of      100 bp
*    40784    41551: contig of 768 bp in length
*    41552 41651: gap of      100 bp
*    41652    42409: contig of 758 bp in length
*    42410 42509: gap of      100 bp
*    42510    43302: contig of 793 bp in length
*    43303 43402: gap of      100 bp
*    43403    44204: contig of 802 bp in length
*    44205 44304: gap of      100 bp
*    44305    45097: contig of 793 bp in length
*    45098 45197: gap of      100 bp
*    45198    45976: contig of 779 bp in length
*    45977 46076: gap of      100 bp
*    46077    46856: contig of 780 bp in length
*    46857 46956: gap of      100 bp
*    46957    47735: contig of 779 bp in length
*    47736 47835: gap of      100 bp
*    47836    48609: contig of 774 bp in length
*    48610 48709: gap of      100 bp
*    48710    49519: contig of 810 bp in length
*    49520 49619: gap of      100 bp
*    49620    50405: contig of 786 bp in length
*    50406 50505: gap of      100 bp
*    50506    51282: contig of 777 bp in length
*    51283 51382: gap of      100 bp
*    51383    52152: contig of 770 bp in length
*    52153 52252: gap of      100 bp
*    52253    53044: contig of 792 bp in length
*    53045 53144: gap of      100 bp
*    53145    53941: contig of 797 bp in length
*    53942 54041: gap of      100 bp
*    54042    54836: contig of 795 bp in length
*    54837 54936: gap of      100 bp
*    54937    55730: contig of 794 bp in length
*    55731 55830: gap of      100 bp
*    55831    56616: contig of 786 bp in length
*    56617 56716: gap of      100 bp
*    56717    57498: contig of 782 bp in length
*    57499 57598: gap of      100 bp
*    57599    58381: contig of 783 bp in length
*    58382 58481: gap of      100 bp
*    58482    59250: contig of 769 bp in length
*    59251 59350: gap of      100 bp
*    59351    60139: contig of 789 bp in length
*    60140 60239: gap of      100 bp
*    60240    61030: contig of 791 bp in length
*    61031 61130: gap of      100 bp
*    61131    61917: contig of 787 bp in length
*    61918 62017: gap of      100 bp
*    62018    62811: contig of 794 bp in length

  Query Match             68.0%;  Score 17;  DB 2;  Length 69900;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           ||||||||||||||||
Db   19825 CGTCAGGTGCAGGTCAG 19809


RESULT  20
AC069132
LOCUS       AC069132   151340 bp    DNA             HTG       17-JUN-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-556L8, *** SEQUENCING IN
            PROGRESS ***, 62 unordered pieces.
ACCESSION   AC069132
VERSION     AC069132.2  GI:8469032
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151340)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151340)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
```

```
COMMENT     On Jun 12, 2000 this sequence version replaced gi:7924001.

            ----------------- Genome Center -----------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 62 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1502: contig of 1502 bp in length
            *     1503     1602: gap of unknown length
            *     1603     2741: contig of 1139 bp in length
            *     2742     2841: gap of unknown length
            *     2842     4036: contig of 1195 bp in length
            *     4037     4136: gap of unknown length
            *     4137     5620: contig of 1484 bp in length
            *     5621     5720: gap of unknown length
            *     5721     7080: contig of 1360 bp in length
            *     7081     7180: gap of unknown length
            *     7181     8839: contig of 1659 bp in length
            *     8840     8939: gap of unknown length
            *     8940     9987: contig of 1048 bp in length
            *     9988    10087: gap of unknown length
            *    10088    11156: contig of 1069 bp in length
            *    11157    11256: gap of unknown length
            *    11257    12404: contig of 1148 bp in length
            *    12405    12504: gap of unknown length
            *    12505    13861: contig of 1357 bp in length
            *    13862    13961: gap of unknown length
            *    13962    15241: contig of 1280 bp in length
            *    15242    15341: gap of unknown length
            *    15342    17009: contig of 1668 bp in length
            *    17010    17109: gap of unknown length
            *    17110    18897: contig of 1788 bp in length
            *    18898    18997: gap of unknown length
            *    18998    20389: contig of 1392 bp in length
            *    20390    20489: gap of unknown length
            *    20490    22042: contig of 1553 bp in length
            *    22043    22142: gap of unknown length
            *    22143    24057: contig of 1915 bp in length
            *    24058    24157: gap of unknown length
            *    24158    26189: contig of 2032 bp in length
            *    26190    26289: gap of unknown length
            *    26290    27959: contig of 1670 bp in length
            *    27960    28059: gap of unknown length
            *    28060    30067: contig of 2008 bp in length
            *    30068    30167: gap of unknown length
            *    30168    32456: contig of 2289 bp in length
            *    32457    32556: gap of unknown length
            *    32557    34597: contig of 2041 bp in length
            *    34598    34697: gap of unknown length
            *    34698    36611: contig of 1914 bp in length
            *    36612    36711: gap of unknown length
            *    36712    38677: contig of 1966 bp in length
            *    38678    38777: gap of unknown length
            *    38778    40974: contig of 2197 bp in length
            *    40975    41074: gap of unknown length
            *    41075    42666: contig of 1592 bp in length
            *    42667    42766: gap of unknown length
            *    42767    44292: contig of 1526 bp in length
            *    44293    44392: gap of unknown length
            *    44393    46032: contig of 1640 bp in length
            *    46033    46132: gap of unknown length
            *    46133    48815: contig of 2683 bp in length
            *    48816    48915: gap of unknown length
            *    48916    51058: contig of 2143 bp in length
            *    51059    51158: gap of unknown length
            *    51159    53094: contig of 1936 bp in length
            *    53095    53194: gap of unknown length
            *    53195    54724: contig of 1530 bp in length
            *    54725    54824: gap of unknown length
            *    54825    57019: contig of 2195 bp in length
            *    57020    57119: gap of unknown length
            *    57120    59736: contig of 2617 bp in length
            *    59737    59836: gap of unknown length
            *    59837    61510: contig of 1674 bp in length
            *    61511    61610: gap of unknown length
            *    61611    63247: contig of 1637 bp in length
            *    63248    63347: gap of unknown length
            *    63348    65262: contig of 1915 bp in length
            *    65263    65362: gap of unknown length
            *    65363    67493: contig of 2131 bp in length
            *    67494    67593: gap of unknown length
            *    67594    70740: contig of 3147 bp in length
            *    70741    70840: gap of unknown length
            *    70841    72256: contig of 1416 bp in length
            *    72257    72356: gap of unknown length
            *    72357    75456: contig of 3100 bp in length
            *    75457    75556: gap of unknown length
            *    75557    77328: contig of 1772 bp in length
            *    77329    77428: gap of unknown length
            *    77429    79851: contig of 2423 bp in length
            *    79852    79951: gap of unknown length
            *    79952    82877: contig of 2926 bp in length
            *    82878    82977: gap of unknown length
            *    82978    84859: contig of 1882 bp in length
            *    84860    84959: gap of unknown length
            *    84960    87856: contig of 2897 bp in length
            *    87857    87956: gap of unknown length
            *    87957    91677: contig of 3721 bp in length
            *    91678    91777: gap of unknown length
            *    91778    94536: contig of 2759 bp in length
            *    94537    94636: gap of unknown length
            *    94637    97377: contig of 2741 bp in length
            *    97378    97477: gap of unknown length
            *    97478   100981: contig of 3504 bp in length
            *   100982   101081: gap of unknown length
            *   101082   103958: contig of 2877 bp in length
            *   103959   104058: gap of unknown length
            *   104059   106903: contig of 2845 bp in length
            *   106904   107003: gap of unknown length
            *   107004   109662: contig of 2659 bp in length
            *   109663   109762: gap of unknown length
            *   109763   113672: contig of 3910 bp in length
            *   113673   113772: gap of unknown length
            *   113773   118353: contig of 4581 bp in length
            *   118354   118453: gap of unknown length
            *   118454   121541: contig of 3088 bp in length
            *   121542   121641: gap of unknown length
            *   121642   124122: contig of 2481 bp in length
            *   124123   124222: gap of unknown length
            *   124223   128114: contig of 3892 bp in length
            *   128115   128214: gap of unknown length
            *   128215   131684: contig of 3470 bp in length
            *   131685   131784: gap of unknown length
            *   131785   135575: contig of 3791 bp in length
            *   135576   135675: gap of unknown length
            *   135676   140587: contig of 4912 bp in length
            *   140588   140687: gap of unknown length
            *   140688   145142: contig of 4455 bp in length
            *   145143   145242: gap of unknown length
            *   145243   151340: contig of 6098 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151340
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-556L8"
BASE COUNT    43058 a  29029 c  29780 g  43356 t   6117 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 151340;
  Best Local Similarity   100.0%;  Pred. No. 19;
```

```
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcag 18
          ||||||||||||||||||
Db  93019 CGTCAGGTGCAGGTCAG 93035


RESULT  21
AC068100/c
LOCUS       AC068100   161742 bp    DNA             HTG       04-JUL-2001
DEFINITION  Homo sapiens chromosome X clone RP11-812N12 map X, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC068100
VERSION     AC068100.5  GI:14595922
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161742)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome X, clone RP11-812N12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161742)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 4, 2001 this sequence version replaced gi:13357469.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7131
                Center clone name: 812_N_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 4% of reads
                Sequencing vector: Plasmid; n/a; 96% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 158467 bases at least Q40
                Consensus quality: 159785 bases at least Q30
                Consensus quality: 160319 bases at least Q20
                Insert size: 165000; agarose-fp
                Insert size: 160742; sum-of-contigs
                Quality coverage: 6.9 in Q20 bases; agarose-fp
                Quality coverage: 7.1 in Q20 ba.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1586: contig of 1586 bp in length
            *     1587 1686: gap of      100 bp
            *     1687     4904: contig of 3218 bp in length
            *     4905 5004: gap of      100 bp
            *     5005     6284: contig of 1280 bp in length
            *     6285 6384: gap of      100 bp
            *     6385    87528: contig of 81144 bp in length
            *    87529 87628: gap of      100 bp
            *    87629    90089: contig of 2461 bp in length
            *    90090 90189: gap of      100 bp
            *    90190    96930: contig of 6741 bp in length
            *    96931 97030: gap of      100 bp
            *    97031   103279: contig of 6249 bp in length
            *   103280 103379: gap of      100 bp
            *   103380   114728: contig of 11349 bp in length
            *   114729 114828: gap of      100 bp
            *   114829   125622: contig of 10794 bp in length
            *   125623 125722: gap of      100 bp
            *   125723   156844: contig of 31122 bp in length
            *   156845 156944: gap of      100 bp
            *   156945   161742: contig of 4798 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161742
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="RP11-812N12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1586
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1687. .4904
                     /note="assembly_fragment"
     misc_feature    5005. .6284
                     /note="assembly_fragment"
     misc_feature    6385. .87528
                     /note="assembly_fragment"
     misc_feature    87629. .90089
                     /note="assembly_fragment"
     misc_feature    90190. .96930
                     /note="assembly_fragment"
     misc_feature    97031. .103279
                     /note="assembly_fragment"
     misc_feature    103380. .114728
                     /note="assembly_fragment"
     misc_feature    114829. .125622
                     /note="assembly_fragment"
     misc_feature    125723. .156844
                     /note="assembly_fragment"
     misc_feature    156945. .161742
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    45576 a  33084 c  34118 g  47963 t   1001 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 161742;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcag 18
          ||||||||||||||||||
Db  70155 CGTCAGGTGCAGGTCAG 70139
```

```
RESULT  22
AC041048/c
LOCUS       AC041048   168360 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-759J14, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC041048
VERSION     AC041048.3  GI:7770701
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168360)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168360)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 12, 2000 this sequence version replaced gi:7622530.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0759J14
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 157456 bases at least Q40
            Consensus quality: 161145 bases at least Q30
            Consensus quality: 163068 bases at least Q20
            Insert size: 173000; agarose-fp
            Insert size: 166760; sum-of-contigs
            Quality coverage: 4.08 in Q20 bases; agarose-fp
            Quality coverage: 4.27 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1711: contig of 1711 bp in length
            *     1712     1811: gap of unknown length
            *     1812     6135: contig of 4324 bp in length
            *     6136     6235: gap of unknown length
            *     6236     8687: contig of 2452 bp in length
            *     8688     8787: gap of unknown length
            *     8788    11647: contig of 2860 bp in length
            *    11648    11747: gap of unknown length
            *    11748    13718: contig of 1971 bp in length
            *    13719    13818: gap of unknown length
            *    13819    18016: contig of 4198 bp in length
            *    18017    18116: gap of unknown length
            *    18117    21570: contig of 3454 bp in length
            *    21571    21670: gap of unknown length
            *    21671    24723: contig of 3053 bp in length
            *    24724    24823: gap of unknown length
            *    24824    32650: contig of 7827 bp in length
            *    32651    32750: gap of unknown length
            *    32751    46491: contig of 13741 bp in length
            *    46492    46591: gap of unknown length
            *    46592    60127: contig of 13536 bp in length
            *    60128    60227: gap of unknown length
            *    60228    74481: contig of 14254 bp in length
            *    74482    74581: gap of unknown length
            *    74582    90886: contig of 16305 bp in length
            *    90887    90986: gap of unknown length
            *    90987   104984: contig of 13998 bp in length
            *   104985   105084: gap of unknown length
            *   105085   127379: contig of 22295 bp in length
            *   127380   127479: gap of unknown length
            *   127480   147028: contig of 19549 bp in length
            *   147029   147128: gap of unknown length
            *   147129   168360: contig of 21232 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168360
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-759J14"
     misc_feature    1. .1711
                     /note="assembly_name:Contig6"
     misc_feature    1812. .6135
                     /note="assembly_name:Contig7"
     misc_feature    6236. .8687
                     /note="assembly_name:Contig8
                     clone_end:SP6
                     vector_side:right"
     misc_feature    8788. .11647
                     /note="assembly_name:Contig9"
     misc_feature    11748. .13718
                     /note="assembly_name:Contig10"
     misc_feature    13819. .18016
                     /note="assembly_name:Contig11"
     misc_feature    18117. .21570
                     /note="assembly_name:Contig12"
     misc_feature    21671. .24723
                     /note="assembly_name:Contig13"
     misc_feature    24824. .32650
                     /note="assembly_name:Contig14"
     misc_feature    32751. .46491
                     /note="assembly_name:Contig15"
     misc_feature    46592. .60127
                     /note="assembly_name:Contig16"
     misc_feature    60228. .74481
                     /note="assembly_name:Contig17"
     misc_feature    74582. .90886
                     /note="assembly_name:Contig18"
     misc_feature    90987. .104984
                     /note="assembly_name:Contig19"
     misc_feature    105085. .127379
                     /note="assembly_name:Contig20"
     misc_feature    127480. .147028
                     /note="assembly_name:Contig21"
     misc_feature    147129. .168360
                     /note="assembly_name:Contig22"
BASE COUNT    48985 a  34329 c  34643 g  48785 t   1618 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 168360;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           ||||||||||||||||||
Db 153451 CGTCAGGTGCAGGTCAG 153435


RESULT  23
AC060833/c
LOCUS       AC060833   176379 bp    DNA             HTG       17-AUG-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-556F3, WORKING DRAFT SEQUENCE,
```

```
            4 unordered pieces.
ACCESSION   AC060833
VERSION     AC060833.3  GI:9838272
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176379)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176379)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 17, 2000 this sequence version replaced gi:8980022.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0556F03
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 173730 bases at least Q40
            Consensus quality: 174579 bases at least Q30
            Consensus quality: 175046 bases at least Q20
            Insert size: 174000; agarose-fp
            Insert size: 177415; sum-of-contigs
            Quality coverage: 6.66 in Q20 bases; agarose-fp
            Quality coverage: 6.55 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1419: contig of 1419 bp in length
            *     1420     1519: gap of unknown length
            *     1520    10016: contig of 8497 bp in length
            *    10017    10116: gap of unknown length
            *    10117    79219: contig of 69103 bp in length
            *    79220    79319: gap of unknown length
            *    79320   176379: contig of 97060 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176379
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-556F3"
     misc_feature    1. .1419
                     /note="assembly_name:Contig5"
     misc_feature    1520. .10016
                     /note="assembly_name:Contig6
                     clone_end:SP6
                     vector_side:right"
     misc_feature    10117. .79219
                     /note="assembly_name:Contig7"
     misc_feature    79320. .176379
                     /note="assembly_name:Contig8
                     clone_end:T7
                     vector_side:left"

BASE COUNT    49569 a  33984 c  35497 g  57025 t    304 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 176379;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           ||||||||||||||||
Db   81954 CGTCAGGTGCAGGTCAG 81938


RESULT  24
AC022781/c
LOCUS       AC022781   220166 bp    DNA             HTG       04-JUL-2001
DEFINITION  Mus musculus chromosome 11 clone RP23-199H17 map 11, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC022781
VERSION     AC022781.6  GI:14595917
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 220166)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 11, clone RP23-199H17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220166)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 4, 2001 this sequence version replaced gi:10305259.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
                Center project name: L5913
                Center clone name: 199_H_17
            ---------- Summary Statistics
                Sequencing vector: M13; M77815; 43% of reads
                Sequencing vector: Plasmid; n/a; 57% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 217289 bases at least Q40
                Consensus quality: 218321 bases at least Q30
                Consensus quality: 218800 bases at least Q20
                Insert size: 206000; agarose-fp
                Insert size: 219466; sum-of-contigs
                Quality coverage: 9.5 in Q20 bases; agarose-fp
```

```
            Quality coverage: 8.9 in Q20 b.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 8 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    22567: contig of 22567 bp in length
          *    22568 22667: gap of      100 bp
          *    22668    28303: contig of 5636 bp in length
          *    28304 28403: gap of      100 bp
          *    28404   109561: contig of 81158 bp in length
          *   109562 109661: gap of      100 bp
          *   109662   118270: contig of 8609 bp in length
          *   118271 118370: gap of      100 bp
          *   118371   140406: contig of 22036 bp in length
          *   140407 140506: gap of      100 bp
          *   140507   158389: contig of 17883 bp in length
          *   158390 158489: gap of      100 bp
          *   158490   211583: contig of 53094 bp in length
          *   211584 211683: gap of      100 bp
          *   211684   220166: contig of 8483 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220166
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /map="11"
                     /clone="RP23-199H17"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .22567
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    22668. .28303
                     /note="assembly_fragment"
     misc_feature    28404. .109561
                     /note="assembly_fragment"
     misc_feature    109662. .118270
                     /note="assembly_fragment"
     misc_feature    118371. .140406
                     /note="assembly_fragment"
     misc_feature    140507. .158389
                     /note="assembly_fragment"
     misc_feature    158490. .211583
                     /note="assembly_fragment"
     misc_feature    211684. .220166
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    53983 a  54774 c  55714 g  54982 t    713 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 220166;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           |||||||||||||||||
Db   37316 CGTCAGGTGCAGGTCAG 37300


RESULT  25
AL591065
LOCUS       AL591065   222212 bp    DNA             HTG       13-JUN-2001
DEFINITION  Mus musculus chromosome 11 clone RP23-199H17, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL591065
VERSION     AL591065.2  GI:13990285
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 222212)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 7, 2001 this sequence version replaced gi:13938846.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM199H17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 218929 bases at least Q40
            Consensus quality: 219403 bases at least Q30
            Consensus quality: 219831 bases at least Q20
            Insert size: 220512; sum-of-contigs
            Insert size: 221854; 2.3% error; agarose-fp
            Quality coverage: 10.86x in Q20 bases; sum-of-contigs Quality
            coverage: 11.12x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .222212
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-199H17"
                     /clone_lib="RPCI-23"
     misc_feature    1. .19515
                     /note="assembly_fragment:03069
                     fragment_chain:1"
     misc_feature    19616. .27961
                     /note="assembly_fragment:00400
                     fragment_chain:1"
     misc_feature    28062. .37082
                     /note="assembly_fragment:03718
                     fragment_chain:1"
     misc_feature    37183. .45022
                     /note="assembly_fragment:06200
                     fragment_chain:1"
     misc_feature    45123. .57322
                     /note="assembly_fragment:00161
                     fragment_chain:1"
     misc_feature    57423. .59552
                     /note="assembly_fragment:02581
                     fragment_chain:1"
     misc_feature    59653. .77089
                     /note="assembly_fragment:00011
                     fragment_chain:1"
     misc_feature    77190. .88794
                     /note="assembly_fragment:05373
                     fragment_chain:1"
     misc_feature    88895. .101172
                     /note="assembly_fragment:05834
                     fragment_chain:1"
     misc_feature    101273. .112983
                     /note="assembly_fragment:02036
                     fragment_chain:1"
     misc_feature    113084. .119048
                     /note="assembly_fragment:00573
```

```
                     fragment_chain:1"
     misc_feature    119149. .139095
                     /note="assembly_fragment:04963
                     fragment_chain:1"
     misc_feature    139196. .141828
                     /note="assembly_fragment:02943
                     fragment_chain:1"
     misc_feature    141929. .150289
                     /note="assembly_fragment:04561
                     fragment_chain:2"
     misc_feature    150390. .166695
                     /note="assembly_fragment:06222
                     fragment_chain:2"
     misc_feature    166796. .169310
                     /note="assembly_fragment:04338
                     fragment_chain:2"
     misc_feature    169411. .210287
                     /note="assembly_fragment:05139
                     fragment_chain:3"
     misc_feature    210388. .222212
                     /note="assembly_fragment:02948
                     fragment_chain:3
                     clone_end:T7
                     vector_side:right"
BASE COUNT    55706 a  56624 c  54341 g  53835 t   1706 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 222212;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcag 18
          |||||||||||||||||
Db 195872 CGTCAGGTGCAGGTCAG 195888


RESULT  26
AF237379
LOCUS       AF237379     8326 bp    DNA             VRT       13-NOV-2000
DEFINITION  Xenopus laevis neurofilament protein (NF-M(2)) gene, complete cds.
ACCESSION   AF237379
VERSION     AF237379.1  GI:11139431
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 8326)
  AUTHORS   Roosa,J.R., Gervasi,C. and Szaro,B.G.
  TITLE     Structure, biological activity of the upstream regulatory sequence,
            and conserved domains of a middle molecular mass neurofilament gene
            of Xenopus laevis
  JOURNAL   Brain Res. Mol. Brain Res. 82 (1-2), 35-51 (2000)
   PUBMED   11042356
REFERENCE   2  (bases 1 to 8326)
  AUTHORS   Roosa,J.R. and Szaro,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2000) Biological Sciences, University at Albany,
            State University of New York, 1400 Washington Avenue, Albany, NY
            12222, USA
FEATURES             Location/Qualifiers
     source          1. .8326
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     TATA_signal     1498. .1503
     5'UTR           1524. .1563
     exon            1524. .2598
                     /number=1
     mRNA            join(1524. .2598,3533. .3657,6507. .>8326)
                     /product="neurofilament protein"
     gene            1564. .>8326
                     /gene="NF-M(2)"
     CDS             join(1564. .2598,3533. .3657,6507. .8088)
                     /gene="NF-M(2)"
                     /codon_start=1
                     /product="neurofilament protein"
                     /protein_id="AAG31697.1"
                     /db_xref="GI:11139432"
                     /translation="MSYSLETIGNPSYRRMAETRTSYSRTSSTPSSGFRSQSWSRSSP
                     TTISSYKRTNLAAPRVYSSNDSLDFSQSSAFNGDLKQVRFSEKEQLQGLNDRFVGYIE
                     KVHFLEQQNKEIEAEIQALRQKQSGYSQLGEVYEQEMRELRVNLELLGHDKAQIVLDT
                     EHLEDDLQRLKERLEEEAIRDETEAAIRAMKKDMEDTSLVKVELDKKVQSLQEEVAF
                     IKNNHEEEVGELLAQIQASSITVERKDFQKTDLASALREIRSQLEGHSSQNMLQTEEV
                     FRNRYAKLTEAAEQNKEAIRSAKDDIAEYRRQLHSKTVELESVRGTKESLERQLNDIE
                     ERHGHDLTSYQETIQQLDNELRGTKWEMSRHLREYQDLLNVKMALDIEIAAYRKLLEG
                     EETRFSSFSGVPTGPSAIHRPTPISSTRVQQARAEPPRVKVQHKFVEEIIEETKVEDD
                     RSDMEAALAAMADEFALGFEEGGEEEEQGEETQEGEEEAEEEIVAAVESSVQAAAPGE
                     EAEGEEEEEEKGEEEAEAEEEGEKEEEEKEEEGEDEGEKVDEEEEGGEQGEEEDVKAE
                     EEEEKEEGEEDAEVEETKAEEGEEEGEKGEDEAEKEEGGEEEEKEEVAADEEGGEGEE
                     EGGEGEKDEDKGEEEKEGEAEAEEEAEGGEEETTEEAQAEEVVEETMTETKIVREKAEV
                     QETEAEEEETQDAEEQKDEGEDEETGEGKETEEKEEGGEEKEDADEGTEQQEEEKAEE
                     EKGEIDEGETKEGAEEDAGEDKQAKSPEEKSSEETISVSKAASVGSEKESAKQEEEKS
                     TQRDDGKEDTAVNGDVEEGEVKEESEDKAVVTNGLDDSPSEDKESSKGQQTVVITKKI
                     ETTTSEGEDGTKYITKSVTVTQKVEEHEETVQEKLVSTKKVEKVTSHAVVKEVTEDK"
     intron          2599. .3532
                     /gene="NF-M(2)"
                     /number=1
     exon            3533. .3657
                     /gene="NF-M(2)"
                     /number=2
     intron          3659. .6507
                     /gene="NF-M(2)"
                     /number=2
     exon            6507. .>8326
                     /gene="NF-M(2)"
                     /number=3
     3'UTR           8089. .>8326
                     /gene="NF-M(2)"
BASE COUNT    2631 a   1560 c   2161 g   1974 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 5;  Length 8326;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagca 20
          ||||||||||||||||
Db   3060 CAGGTGCAGGTCAGCA 3075


RESULT  27
AE004050
LOCUS       AE004050    10536 bp    DNA             BCT       15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 196 of 229 of the complete genome.
ACCESSION   AE004050 AE003849
VERSION     AE004050.1  GI:9107594
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 10536)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
```

TITLE The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis
JOURNAL Nature 406 (6792), 151-157 (2000)
MEDLINE 20365717
PUBMED 10910347
REFERENCE 2 (bases 1 to 10536)
AUTHORS Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S., Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S., Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M., Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

FEATURES Location/Qualifiers

source 1. .10536
/organism="Xylella fastidiosa 9a5c"
/db_xref="taxon:160492"
/clone="9a5c"

gene 1296. .2339
/gene="XF2409"

CDS 1296. .2339
/gene="XF2409"
/note="similar to SP|P39369 (percent identity: 47 %/query alignment coverage: 96.8 %/subject alignment coverage: 99.4 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/Start codon shift: 2523"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF85208.1"
/db_xref="GI:9107595"
/translation="MPALGYLHVDGLCQKSSGGSRYNLYEAETIAAWLAEHHESLQAQ
YGEPLHRIVGILTPFETQARAISQACRDVGIEVGHEKDGITVGTVHALQGAERPVVIF
SAVYSKHADGGFIDQRASMLNVAVSRAKNTFLVFGDMDVFTAAPKSRPRGLLAHYLFK
DPSNALCFQPLVRKDLQQVSTAVEVLQDAAEHDAFLLQALNKVQREIHIVSPWINKDR
IQDIGAFKAMQEAVERQVQVTVYTDQDLNTDDKKDTKKIAEVLQAARALRGVGIEVHF
VDRVHSKMVIGDDEVFCVGSFNWFSANRSAMYAKHETSLVYRGRGLADERQTRLNSLR
QRITDDPQQVGQA"

gene 2311. .2484
/gene="XF2410"

CDS 2311. .2484
/gene="XF2410"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85209.1"
/db_xref="GI:9107596"
/translation="MILSKSARRDAVLRQRDVDTGWRQAMHAAAEGRDMVSPAVALRA
VALLKLCCCLPLG"

gene complement(3181. .5514)
/gene="XF2411"

CDS complement(3181. .5514)
/gene="XF2411"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85210.1"
/db_xref="GI:9107597"
/translation="MLKKLIKVNHVGLLRNGVPPTTPTPFGPVTLFYAENGRGKTTVA
NILRAVTHGKDKDEVLQKARTIDTDDPPCIELLWNDGNTDIKVTLHHGHWTGTPPDIL
IFDATFVDDNVYSGQEVHPAHRQKFLQFVLGAESVKLEREIVRLAKEIADEHRNIKER
EQRIRAYSKNMPIADFIRLIPVTDAEAKIAALRTQREAVRNAPALLQRHAPQPLPHIE
CNIDAFFSIIQSTFHDIRADAEEIVQAHFQRHQQPPGIEAWVMQGREFNAEQGCPFCG
QEVTATALIKAYETYFNTAFKAFMKNVATLGSGVATRFPEACLEKIADRIQGNDACVA
AWKDQLDLSSPFFDNPRATATFTRLRTVALALAQSKQHRPLDCVGTTQDKADMVRLLE
ELRAQMDAYNAAVVEINRRITNFRKTLAENTPAALDASIATLEIGIVRQSTEVVQAIT
VYQAAKAKKEQLEREKKNMRAALDAGLPDLLSEYASKINQFLRDFGAAFSIKKLQQSM
QGGTMRASYVLQLRGKEVALGRRTDSDPGFHNVLSEGDKRTLALAFFLARLYATPDAL
VGKSVVLDDPMCSFDMTRRNRTMESIAALVNQGVQVVVLSHDAYFLRDLRDLLAGPRY
NKVSVNVHHIKRTKNNDSQIVSDADLDTICQSPYMLRYAQVVTFVSGTYEGTLQEVAS
ALRPLVEGFLKHRFAPPLLRQDLSLGQMISDIRKATDDSPLVLAKPYVDTLEKLNAFA
VQSHHDDSKSLSPMNDGQLRQYARMALELIYGGSLPH"

gene complement(5615. .5800)
/gene="XF2412"

CDS complement(5615. .5800)
/gene="XF2412"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85211.1"
/db_xref="GI:9107598"
/translation="MIFGATQTNADRRRDSGTRHAVCTGWARSRLPPDPQVRHNAHLP
KPHVKVISSIPITGTML"

gene 6084. .6512
/gene="XF2413"

CDS 6084. .6512
/gene="XF2413"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85212.1"
/db_xref="GI:9107599"
/translation="MVGGDVPTADTRAAAGGACEPSIKNIALSRGVHVLLLCAGMRLW
PGQGCLGQNGPMKGSVVKSSVVLGVVLRCGSQCQQVEAAARGFAPRTFGKYCMSALWL
YRRWCSERHCLSCRTMEGMGATSGDAGPRRGRGWLGKGRG"

gene complement(6509. .8314)
/gene="XF2414"

CDS complement(6509. .8314)
/gene="XF2414"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF85213.1"
/db_xref="GI:9107600"
/translation="MPPKRHQVHDAADTDKENLTVTTEHCFIGVCFIPLDGSGQTFFG
FSEFLAGLALMVLAWTIADVRYRFRIQVAPLPLLRMSFYVVAAIGLLTLLTDLWRAEQ
WLVPRGILLTPAVWQALLGGAFLLTILTWAWLGFIRPPTYGQYNAARFQQTLYRIILN
GSPTELAVIADELVHSAEALVQHATNRYSLQNGERIECEEEEQEDLPQVNRDANALLL
LIAEKRLCRAIVAESPRTAFALFQAIANTKKYGIPIRVFAQNIFHEALANKDSFLYHE
TNGCALGLIGDHKPFTQAMFANYEMVETIGTLFNHDRDARNKWDFVQWLAYCSVVLMT
LRSYAKQGYRGYSFVLSQAMGYIASAASDLSTLNEPSSAADSDVLRRLALVIQFIKDA
VRILDQEGVPEHVCICVRAGYDSSRAIFCDRIANMIFTVISLASEVRLPQLEWKIQHN

```
                     SVWSNLFHYGAFKRPAGQVVAFKLRRLLYNQIADMQRFPNVERAKILRFCLNVLALAV
                     IRGNDYKSDRRFHHRVLHRAVLAWTKKHFVWLHSHHPRLAESCLGNHITYDAEHLRLA
                     YTHPAKGPIGKVSYFALNPASPPPEPVAASAVPNA"
     tRNA            complement(8365. .8440)
                     /gene="tRNA-Thr-2"
                     /note="found by tRNAscan"
                     /product="tRNA-Thr"
     gene            complement(8365. .8440)
                     /gene="tRNA-Thr-2"
     gene            complement(8457. .8648)
                     /gene="XF2415"
     CDS             complement(8457. .8648)
                     /gene="XF2415"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF85214.1"
                     /db_xref="GI:9107601"
                     /translation="MPVPRAVDPPQPPILKSHLCATVAHLLQSIPSPDPALILMRRAP
                     EAHTKQSSALSRRPSTAMM"
     gene            complement(8921. .9871)
                     /gene="XF2416"
     CDS             complement(8921. .9871)
                     /gene="XF2416"
                     /note="similar to GI|5915671 (percent identity: 64 %/query
                     alignment coverage:  99.4 %/subject alignment coverage:
                     99.4 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="drug tolerance protein"

  Query Match             64.0%;  Score 16;  DB 1;  Length 10536;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtca 17
          ||||||||||||||||
Db   1995 CGTCAGGTGCAGGTCA 2010


RESULT  28
AE005241
LOCUS       AE005241    12242 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 65
            of 155.
ACCESSION   AE005241 AE005174
VERSION     AE005241.1  GI:12513509
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 12242)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 12242)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .12242
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            complement(146. .952)
                     /gene="rna"
                     /note="Z0755"
     CDS             complement(146. .952)
                     /gene="rna"
                     /function="enzyme; Macromolecule degradation: Degradation
                     of RNA"
                     /note="Residues 1 to 268 of 268 are 99.25 pct identical to
                     residues 1 to 268 of 268 from Escherichia coli K-12 Strain
                     MG1655: B0611"
                     /codon_start=1
                     /transl_table=11
                     /product="RNase I, cleaves phosphodiester bond between any
                     two nucleotides"
                     /protein_id="AAG54946.1"
                     /db_xref="GI:12513510"
                     /translation="MKAFWRNAALLAVSLLPFSSANAVALQAKQYGDFDRYVLALSWQ
                     TGFCQSQHDRNRNERDECRLQTETTNKADFLTVHGLWPXLPKSVAARGVDERRWMRFG
                     CATRPIPNLPEARASRMCSSPETGLSLETAAKLSEVMPGAGGRSCLERYEYAKHGACF
                     GFDPDAYFGTMVRLNQEIKESEAGKFLADNYGKTVSRRDFDAAFAKSWGKENVKAVKL
                     TCQGNPAYLTEIQVSIKADAINAPLSANSFLPQPHPGNCGKTFVIDKAGY"
     gene            complement(1066. .2529)
                     /gene="ybdS"
                     /note="Z0756"
     CDS             complement(1066. .2529)
                     /gene="ybdS"
                     /function="putative membrane; Not classified"
                     /note="Residues 1 to 487 of 487 are 100.00 pct identical
                     to residues 1 to 487 of 487 from Escherichia coli K-12
                     Strain MG1655: B0612"
                     /codon_start=1
                     /transl_table=11
                     /product="putative a membrane protein"
                     /protein_id="AAG54947.1"
                     /db_xref="GI:12513511"
                     /translation="MSLAKDNIWKLLAPLVVMGVMFLIPVPDGMPPQAWHYFAVFVAM
                     IVGMILEPIPATAISFIAVTICVIGSNYLLFDAKELADPAFNAQKQALKWGLAGFSST
                     TVWLVFGAFIFALGYEVSGLGRRIALFLVKFMGKRTLTLGYAIVIIDILLAPFTPSNT
                     ARTGGTVFPVIKNLPPLFKSFPNDPSARRIGGYLMWMMVISTSLSSSMFVTGAAPNVL
                     GLEFVSKIAGIQISWLQWFLCFLPVGVILLIIAPWLSYVLYKPEITHSEEVATWAGDE
                     LKTMGALTRREWTLIGLVLLSLGLWVFGSEVINATAVGLLAVSLMLALHVVPWKDITR
                     YNSAWNTLVNLATLVVMANGLTRSGFIDWFAGTMSTHLEGFSPNATVIVLVLVFYFAH
                     YLFASLSAHTATMLPVILAVGKXIPGVPMEQLCILLVLSIGIMGCLTPYATGPGVIIY
                     GCGYVKSKDYWRLGAIFGVIYISMLLLVGWPILAMWN"
     gene            complement(2580. .3458)
                     /gene="citG"
                     /note="Z0757"
     CDS             complement(2580. .3458)
                     /gene="citG"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 292 of 292 are 98.97 pct identical to
                     residues 1 to 292 of 292 from Escherichia coli K-12 Strain
                     MG1655: B0613"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG54948.1"
                     /db_xref="GI:12513512"
                     /translation="MSMPATLTKTTKLATSLIDEYALLGWRAMLTEVNLSPKPGLVDR
                     INCGAHKDMALEDFHRSALAIQGWLPRFIEFGACSAEMAPEAVLHGLRPIGMACEGDM
                     FRATAGVNTHKGSIFSLGLLCAAIGRLLQLNQLLTPTTVCSTAASFCRGLTDRELRTN
                     NSQLTAGQRLYQQLGLTGARGEAEAGYPLVINHALPHYLTLLDQGLDPELALLDTLLL
                     LMAINGDTNVASRGGEGGLRWLQREAQTLLQKGGIRTPADLDYLRQFDRECIERNLSP
```

```
                GGSADLLILTWFLAQI"
gene            complement(3433. .3984)
                /gene="ybdU"
                /note="Z0758"
CDS             complement(3433. .3984)
                /gene="ybdU"
                /function="orf; Unknown function"
                /note="Residues 1 to 183 of 183 are 99.45 pct identical to
                residues 1 to 183 of 183 from Escherichia coli K-12 Strain
                MG1655: B0614"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAG54949.1"
                /db_xref="GI:12513513"
                /translation="MHLLPELASHHAVSIPELLVSRDERQARQHVWLKRHPVPLVSFT
                VVAPGPIKDSEVTRRIFNHGVTALRALAAKQGWQIQEQAALVSASGPEGMLSIAAPAR
                DLKLATIELEHXHPLGRLWDIDVLTPEGEILSRRDYSLPPRRCLLCEQSAAVCARGKT
                HQLTDLLNRMEALLNDVDACNVN"
gene            complement(3988. .5520)
                /gene="citF"
                /note="Z0759"
CDS             complement(3988. .5520)
                /gene="citF"
                /function="enzyme; Central intermediary metabolism: Pool,
                multipurpose conversions of intermed. met'm"
                /note="Residues 1 to 510 of 510 are 99.41 pct identical to
                residues 1 to 510 of 510 from Escherichia coli K-12 Strain
                MG1655: B0615"
                /codon_start=1
                /transl_table=11
                /product="citrate lyase alpha chain"
                /protein_id="AAG54950.1"
                /db_xref="GI:12513514"
                /translation="MTQKIEQSQRQERVAAWNRRAECDLAAFQNSPKQTYQAEKARDR
                KLCANLEEAIRRSGLQDGMTVSFHHAFRGGDLTVNMVMDVIAKMGFKNLILASSSLSD
                CHAPLVEHIRQGVVTRIYTSGLRGPLAEEISRGLLAEPVQIHSHGGRVHLVQSGELNI
                DVAFLGVPSCDEFGNANGYTGKACCGSLGYAMVDADNAKQVVMLTEELLPYPHNPASI
                EQDQVDLIVKVDRVGDAAKIGAGATRMTTNPRELLIARSAADVIVNSGYFKEGFSMQT
                GTGGASLAVTLFLEDKMRSRDIRADFALGGITATMVDLHEKGLIRKLLDVQSFDSHAA
                QSLARNPNHIEISANQYANWGSKGASVDRLDVVVLSALEIDTQFNVNVLTGSDGVLRG
                ASGGHCDTAIASALSIIVAPLVRGRIPTLVDNVLTCITPGSSVDILVTDHGIAVNPAR
                PELAERLQEAGIKVVSIEWLRERARLLTGEPQPIEFTDRVVAVVRYRDGSVIDVVHQV
                KE"
gene            complement(5531. .6454)
                /gene="citE"
                /note="Z0760"
CDS             complement(5531. .6454)
                /gene="citE"
                /function="enzyme; Central intermediary metabolism: Pool,
                multipurpose conversions of intermed. met'm"
                /note="Residues 1 to 307 of 307 are 99.34 pct identical to
                residues 1 to 307 of 307 from Escherichia coli K-12 Strain
                MG1655: B0616"
                /codon_start=1
                /transl_table=11
                /product="citrate lyase beta chain (acyl lyase subunit)"
                /protein_id="AAG54951.1"
                /db_xref="GI:12513515"
                /translation="MGGLPMISASLQQRKTRTRRSMLFVPGANAAMVSNSFIYPADAL
                MFDLEESVALREKDTARRMVYHALQHPLYRDIETIVRVNALDSEWGVNDLEAVVRGGA
                DVVRLPKTDTAQDVLDIEKEILRIEKACGREPGSTGLLAAIESPLGITRAVEIAHASE
                RLIGIALGAEDYVRNLRTERSPEGTELLFARCSILQAARSAGIQAFDTVYSDANNEAG
                FLQEAAHIKQLGFDGKSLINPRQIDLLHNLYAPTQKEVDHARRVVEAAEAAAREGLGV
                VSLNGKMVDGPVIDRARLVXSRAELSGIREE"
gene            complement(6436. .6732)
                /gene="citD"
                /note="Z0761"
CDS             complement(6436. .6732)
                /gene="citD"
                /function="enzyme; Central intermediary metabolism: Pool,
                multipurpose conversions of intermed. met'm"
                /note="Residues 1 to 98 of 98 are 100.00 pct identical to
                residues 1 to 98 of 98 from Escherichia coli K-12 Strain
                MG1655: B0617"
                /codon_start=1
                /transl_table=11
                /product="citrate lyase acyl carrier protein (gamma
                chain)"
                /protein_id="AAG54952.1"
                /db_xref="GI:12513516"
                /translation="MKINQPAVAGTLESGDVMIRIAPLDTQDIDLQINSSVEKQFGDA
                IRTTILDVLARYNVRGVQLNVDDKGALDCILRARLEALLARASGIPALPWEDCQ"
gene            complement(6747. .7892)
                /gene="citC"
                /note="Z0762"
CDS             complement(6747. .7892)
                /gene="citC"
                /function="enzyme; Central intermediary metabolism: Pool,
                multipurpose conversions of intermed. met'm"
                /note="Residues 1 to 381 of 381 are 99.21 pct identical to
                residues 1 to 381 of 381 from Escherichia coli K-12 Strain
                MG1655: B0618"
                /codon_start=1
                /transl_table=11
                /product="citrate lyase synthetase (citrate (pro-3S)-lyase
                ligase"
                /protein_id="AAG54953.1"
                /db_xref="GI:12513517"

  Query Match             64.0%;  Score 16;  DB 1;  Length 12242;
  Best Local Similarity   100.0%;  Pred. No. 97;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 tgcaggtcagcacgtt 24
           ||||||||||||||||
Db    4236 TGCAGGTCAGCACGTT 4251


RESULT  29
MLCB1450
LOCUS       MLCB1450    38065 bp    DNA             BCT       27-AUG-1999
DEFINITION  Mycobacterium leprae cosmid B1450.
ACCESSION   AL035159
VERSION     AL035159.1  GI:4154036
KEYWORDS    60 kD chaperonin; CDP-diacylglycerol--serine
            O-phosphatidyltransferase; cpn60; glutamine-binding protein;
            groEL-2; hydrolase; oxidoreductase; phosphomethylpyrimidine kinase;
            pseudogene; pssA; RLEP1; serine/threonine protein kinase; thiamine
            biosynthesis protein ThiG; thiamine-phosphate pyrophosphorylase;
            thiD; thiE; thiG.
SOURCE      Mycobacterium leprae.
  ORGANISM  Mycobacterium leprae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1  (bases 1 to 38065)
  AUTHORS   Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
  TITLE     Use of an ordered cosmid library to deduce the genomic organization
            of Mycobacterium leprae
  JOURNAL   Mol. Microbiol. 7 (2), 197-206 (1993)
  MEDLINE   93188700
REFERENCE   2  (bases 1 to 38065)
  AUTHORS   Brown,D. and Churcher,C.M.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 38065)
  AUTHORS   James,K.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JAN-1998) Mycobacterium leprae sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr.
            Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne,
            Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,
            France Requests for cosmids should be sent to Karin Eiglmeier
            (kei@pasteur.fr)
COMMENT     Notes:
```

The Sanger Centre is funded to complete the sequence of M. leprae by the Heiser Program for Research in Leprosy and Tuberculosis of The New York Community Trust.
Work in Paris is supported by the Heiser Trust, the Association Francaise Raoul Follereau and the Groupement de Recherches et des Etudes des Genomes (GIP-GREG).
Details of M. leprae sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/)
CDS are numbered using the following system eg MLCB33.01c. ML (M. leprae), cB33 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. All CDS over 100 codons have been analysed. Gene prediction is based on positional base preference in codons especially where there is an increase in the observed/expected third position G + C. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

FEATURES Location/Qualifiers

source 1. .38065
/organism="Mycobacterium leprae"
/db_xref="taxon:1769"
/clone="cosmid B1450"

gene complement(1. .284)
/gene="MLCB1450.01c"

CDS complement(<1. .284)
/gene="MLCB1450.01c"
/note="MLCB1450.01c, probable transcriptional regulator, partial CDS, len: >94 aa; similar to e.g. TR:Q50887 (EMBL:Z56280) Myxococcus xanthus transcription factor CARD (316 aa), fasta scores; opt: 190 z-score: 372.9 E(): 1.8e-13, 28.6% identity in 98 aa overlap. Equivalent to M.tuberculosis Rv3583c (97.9% identity in 94 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAA22685.1"
/db_xref="GI:4154037"
/db_xref="SPTREMBL:Q9ZBM8"
/translation="MIFKVGDTVVYPHHGAALVEAIETRTINGEQKEYLVLKVAQGDL TVRVPAENAEYVGVRDVVGQEGLDQVFQVLRAPHTEEPTNWSRRYKANLE"

gene 654. .1205
/gene="MLCB1450.02"

CDS 654. .1205
/gene="MLCB1450.02"
/note="MLCB1450.02, probable lipoprotein, len: 183 aa; Equivalent to M.tuberculosis Rv3584 (MTV024.02), 63.4% identity in 175 aa overlap. Contains N-terminal signal sequence with appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="putative lipoprotein"
/protein_id="CAA22686.1"
/db_xref="GI:4154038"
/db_xref="SPTREMBL:Q9ZBM7"
/translation="MSRFKISLPALATRVAVLGFLTLMASVLGGCGAGQISQTATQEP AVNGNRVTLNNLALRDIRIQAAQTGDFLQSGRTVDLMLVA1NNSPYVTDRLVSITSDI GTVALNGYTQLPTNGMLFIGTSEGQRIKPPPLQSNNIAKAIVTLAKPITNGLTYNFTF NFEKAGQANVAVPVSAGLAPRQT"

misc_feature 714. .746
/gene="MLCB1450.02"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"

gene 1428. .2463
/gene="radA"

CDS 1428. .2463
/gene="radA"
/note="MLCB1450.03, probable radA pseudogene, len: 1033 bp, Equivalent to M.tuberculosis Rv3585, (MTV024.03), probable DNA repair protein RADA homolog"
/codon_start=1
/pseudo
/transl_table=11
/product="putative radA pseudogene"
/db_xref="PID:e1370359"

misc_feature complement(1723. .5324)
/note="identical to EMBL:MLANTM M14341 M.leprae 65 kd antigen, from 12 to 3613"

repeat_unit complement(2559. .3208)
/note="RLEP1"

gene 2832. .3170
/gene="MLCB1450.04"

CDS 2832. .3170
/gene="MLCB1450.04"
/note="MLCB1450.04, CDS in RLEP element, len: 112 aa; identical to CDS in other RLEP elements e.g. TR:O32969 (MLCB22.36C)"
/codon_start=1
/transl_table=11
/product="hypothetical protein MLCB1450.04"
/protein_id="CAA22688.1"
/db_xref="GI:4154039"
/translation="MTTPTPQGHDMHTKTPLPRGANNYPHTHACIDIAFSTAQVPSPW HHQHVDQAAPTTDMLTCAALIVSTAAKHTKPHRKQAVSHPPTKTPQHSKTRQQDSGLP PTTTEHIGQS"

gene complement(3504. .5129)
/gene="groEL-2"

CDS complement(3504. .5129)
/gene="groEL-2"
/note="MLCB1450.05c, groEL-2, 60 kD chaperonin, len: 541 aa; identical to CH62_MYCLE, and highly similar to many e.g. CH60_BACSU 60 KD chaperonin (543 aa), fasta scores; opt: 2217 z-score: 2520.3 E(): 0, 63.3% identity in 539 aa overlap. Equivalent to M.tuberculosis CH62_MYCTU Rv0440 (94.8% identity in 541 aa overlap). Contains PS00296 Chaperonins cpn60 signature and Pfam match to entry PF00118 cpn60_TCP1, TCP-1 /cpn60 chaperonin family"
/codon_start=1
/transl_table=11
/product="60 kD chaperonin"
/protein_id="CAA22689.1"
/db_xref="GI:4154040"
/translation="MAKTIAYDEEARRGLERGLNSLADAVKVTLGPKGRNVVLEKKWG APTITNDGVSIAKEIELEDPYEKIGAELVKEVAKKTDDVAGDGTTTATVLAQALVKEG LRNVAAGANPLGLKRGIEKAVDKVTETLLKDAKEVETKEQIAATAAISAGDQSIGDLI AEAMDKVGNEGVITVEESNTFGLQLELTEGMRFDKGYISGYFVTDAERQEAVLEEPYI LLVSSKVSTVKDLLPLLEKVIQAGKSLLIIAEDVEGEALSTLVVNKIRGTFKSVAVKA PGFGDRRKAMLQDMAILTGAQVISEEVGLTLENTDLSLLGKARKVVMTKDETTIVEGA GDTDAIAGRVAQIRTEIENSDSDYDREKLQERLAKLAGGVAVIKAGAATEVELKERKH RIEDAVRNAKAAVEEGIVAGGGVTLLQAAPALDKLKLTGDEATGANIVKVALEAPLKQ IAFNSGMEPGVVAEKVRNLSVGHGLNAATGEYEDLLKAGVADPVKVTRSALQNAASIA GLFLTTEAVVADKPEKTAAPASDPTGGMGGMDF"

misc_feature complement(3564. .5066)
/gene="groEL-2"
/note="Pfam match to entry PF00118 cpn60_TCP1, TCP-1 /cpn60 chaperonin family, score 778.80, E-value 2.2e-230"

misc_feature complement(3888. .3923)
/gene="groEL-2"
/note="PS00296 Chaperonins cpn60 signature"

gene complement(5338. .5907)
/gene="MLCB1450.06c"

CDS complement(5338. .5907)

```
                     /gene="MLCB1450.06c"
                     /note="MLCB1450.06c, probable transcriptional regulator,
                     len: 189 aa; similar to hypothetical proteins from M.
                     tuberculosis and S. coelicolor, and equivalent to
                     M.tuberculosis putative transcriptional regulator Rv0067c
                     (68.8% identity in 189 aa overlap). Contains probable
                     helix-turn-helix motif at aa 35-56 (Score 1794, +5.30 SD)
                     and Pfam match to entry PF00440 tetR, Bacterial regulatory
                     proteins, tetR family"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcriptional regulator"
                     /protein_id="CAA22690.1"
                     /db_xref="GI:4154041"
                     /db_xref="SPTREMBL:Q9ZBM6"
                     /translation="MAQPARPLRADAARNRARILGVAYEAFATEGLSVPIDEIARRAG
                     VGAGTVYRHFPTKEALCAAVIGDRMPHLVDDGYVLLKSAGPGEALFTYLRSLVLHWGA
                     TDRGLVDALAGAGGIDILGVAPDAEDAFLTILSDLLYAAQYAGTARTDVGVREVKSIL
                     VGCQAMEAYNSALAERVTDVVVDGLRAAR"

  Query Match             64.0%;  Score 16;  DB 1;  Length 38065;
  Best Local Similarity   100.0%;  Pred. No. 80;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   14193 CAGGTGCAGGTCAGCA 14208


RESULT  30
AC091120
LOCUS       AC091120    39520 bp    DNA              HTG       30-MAR-2001
DEFINITION  Leishmania major chromosome 2 clone L6541 strain Friedlin, ***
            SEQUENCING IN PROGRESS ***, 3 ordered pieces.
ACCESSION   AC091120
VERSION     AC091120.1  GI:13489160
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Leishmania major.
  ORGANISM  Leishmania major
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Leishmania.
REFERENCE   1  (bases 1 to 39520)
  AUTHORS   Myler,P.J., Sisk,E., Ruiz,J., Cosenza,P., Cruz,A., Stuart,K.,
            Robertson,L., Vogt,C., Marty,A., Cawthra,J., Seyler,A., Munden,H.,
            Nelson,S., Mack,J. and Pentony,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2001) Seattle Biomedical Research Institution, 4
            Nickerson Street, Seattle, WA 98109-1651, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    27709: contig of 27709 bp in length
            *                    gap of unknown length
            *    27710    38388: contig of 10679 bp in length
            *                    gap of unknown length
            *    38389    39520: contig of 1132 bp in length.
FEATURES             Location/Qualifiers
     source          1. .39520
                     /organism="Leishmania major"
                     /strain="Friedlin"
                     /db_xref="taxon:5664"
                     /chromosome="2"
                     /clone="L6541"
BASE COUNT     6965 a  12678 c  12387 g   7487 t      3 others
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 2;  Length 39520;
  Best Local Similarity   100.0%;  Pred. No. 80;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   1027 CAGGTGCAGGTCAGCA 1042


RESULT  31
AC004508
LOCUS       AC004508    71230 bp    DNA              PRI       30-MAR-1998
DEFINITION  Homo sapiens chromosome 21, P1 clone LBL#3 (LBNL H09), complete
            sequence.
ACCESSION   AC004508 L35684 L35685 L35686 L35687 AC000939 AC000938 AC000940
            AC000937 L81573 L35689 L35690 L35691 L35692 L35693 L35694 L35695
            L35696 L35697 L35698 L35699 L42077 L35700 L35701
VERSION     AC004508.1  GI:2996630
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 71230)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Sequencing of human chromosome 21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71230)
  AUTHORS   Ricke,D.O.
  TITLE     Large Scale Sequence Analysis and Annotation with the Sequence
            Comparison Analysis (SCAN) System
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 71230)
  AUTHORS   Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
            Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
            Rojeski,H., Subramanian,S. and Martin,C.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
            DOE Joint Genome Institute.
FEATURES             Location/Qualifiers
     source          1. .71230
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="LBL#3"
                     /note="LBNL H09"
     repeat_region   complement(472. .788)
                     /rpt_family="MLT1a"
     repeat_region   1408. .1743
                     /rpt_family="Alu"
     misc_feature    4057. .4515
                     /note="99% identity zh62d01.r1"
                     /db_xref="dbEST:W86547"
     misc_feature    complement(4506. .4622)
                     /note="GRAIL 2 excellent exon, frame 2"
     misc_feature    complement(4646. .5114)
                     /note="100% identity zh62d01.s1"
                     /db_xref="dbEST:W86451"
     STS             complement(4646. .5114)
                     /standard_name="SHGC-52452"
                     /note="100% identity"
                     /db_xref="dbSTS:G33232"
     repeat_region   complement(5148. .5285)
                     /rpt_family="Alu"
     repeat_region   complement(5788. .5902)
                     /rpt_family="MER44C"
     repeat_region   complement(5974. .6182)
                     /rpt_family="MER44C"
```

```
     repeat_region    complement(6153. .6473)
                      /rpt_family="Alu"
     repeat_region    complement(6485. .6565)
                      /rpt_family="MER44C"
     repeat_region    complement(6779. .6989)
                      /rpt_family="MER46"
     repeat_region    complement(7235. .8312)
                      /rpt_family="L1"
     repeat_region    complement(8410. .8569)
                      /rpt_family="L1"
     repeat_region    8577. .8987
                      /rpt_family="THE1"
     repeat_region    complement(9020. .9511)
                      /rpt_family="L1"
     repeat_region    9872. .10296
                      /rpt_family="THE1"
     repeat_region    10297. .11909
                      /rpt_family="MSTAR"
     repeat_region    11929. .12309
                      /rpt_family="THE1"
     repeat_region    12868. .13131
                      /rpt_family="Alu"
     repeat_region    complement(13267. .13635)
                      /rpt_family="L1"
     prim_transcript  join(14992. .15122,21091. .21151,22200. .22263,25113. .25251,
                      33644. .33721)
                      /standard_name="yj99d04"
                      /note="98% identity R74138"
     misc_feature     14993. .15116
                      /note="GRAIL 2 excellent exon, frame 0"
     prim_transcript  join(15038. .15122,21091. .21151,22209. .22263,25113. .25251,
                      33644. .33885)
                      /standard_name="zu55b03"
                      /note="98% identity AA402132"
     repeat_region    complement(15991. .16164)
                      /rpt_family="MER42"
     repeat_region    complement(16400. .16743)
                      /rpt_family="MLT2B2"
     repeat_region    complement(16761. .17073)
                      /rpt_family="Alu"
     repeat_region    16768. .16787
                      /note="(T)20"
                      /rpt_type=tandem
                      /rpt_unit=T
     misc_feature     19581. .19626
                      /note="GRAIL 2 excellent exon, frame 0"
     repeat_region    20435. .20765
                      /rpt_family="Alu"
     repeat_region    23936. .24290
                      /rpt_family="Alu"
     repeat_region    24273. .24292
                      /note="(A)20"
                      /rpt_type=tandem
                      /rpt_unit=A
     misc_feature     25115. .25247
                      /note="GRAIL 2 excellent exon, frame 2"
     repeat_region    26872. .26965
                      /rpt_family="MIR"
     repeat_region    29019. .29360
                      /rpt_family="Alu"
     repeat_region    29356. .31388
                      /rpt_family="L1"
     repeat_region    31440. .32072
                      /rpt_family="L1"
     repeat_region    complement(32120. .32230)
                      /rpt_family="L1"
     repeat_region    32581. .32624
                      /note="(AC)22"
                      /rpt_type=tandem
                      /rpt_unit=AC
     repeat_region    complement(32668. .32776)
                      /rpt_family="MER3"
     misc_feature     32932. .33120
                      /note="86% identity zh52f10.r1"
                      /db_xref="dbEST:W78940"
     repeat_region    complement(33133. .33199)
                      /rpt_family="MER3"
     repeat_region    complement(33644. .33883)
                      /note="99% identity nf66d01.s1"
                      /db_xref="dbEST:AA515600"
     misc_feature     33648. .33737
                      /note="GRAIL 2 excellent exon, frame 2"
     repeat_region    35260. .35660
                      /rpt_family="MLT1"
     repeat_region    36246. .36533
                      /rpt_family="Alu"
     repeat_region    complement(39299. .39589)
                      /rpt_family="Alu"
     repeat_region    complement(39781. .39931)
                      /rpt_family="Alu"
     repeat_region    complement(41134. .41588)
                      /rpt_family="MLT2"
     repeat_region    41621. .41706
                      /rpt_family="MLT1"
     repeat_region    41748. .41771
                      /note="(GTTT)6"
                      /rpt_type=tandem
                      /rpt_unit=GTTT
     repeat_region    complement(41757. .42397)
                      /rpt_family="L1"
     repeat_region    42935. .43106
                      /rpt_family="Alu"
     misc_feature     complement(44760. .44802)
                      /note="GRAIL 2 excellent exon, frame 2"
     repeat_region    complement(49449. .49589)
                      /rpt_family="MIR2"
     repeat_region    complement(51869. .52337)
                      /rpt_family="MLT1"
     repeat_region    52773. .53082
                      /rpt_family="Alu"
     gene             <53503. .>67215
                      /gene="MX2"
                      /note="interferon-regulated resistance GTP-binding protein
                      MXB (P78-related protein)"
                      /evidence=experimental
     CDS              join(<53503. .53668,54302. .54494,56530. .56664,58923. .59077,
                      67077. .>67215)
                      /gene="MX2"
                      /note="Interferon-regulated resistance GTP-binding protein
                      MXB (P78-related protein)"
                      /codon_start=2
                      /evidence=experimental
                      /product="MX2"
                      /protein_id="AAC08448.1"
                      /db_xref="GI:2996631"
                      /translation="FQQQPPPFGTVPPQMMFPPNWQGAEKDAAFLAKDFNFLTLNNQP
                      PPGNRSQPRAMGPENNLYSQYEQKVRPCIDLIDSLRALGVEQDLALPAIAVIGDQSSG
                      KSSVLEALSGVALPRGSGIVTRCPLVLKLKKQPCEAWAGRISYRNTELELQDPGQVEK
                      EIHKAQNVMAGNGRGISHELISLEITSPEVPDLTIIDLPGITRVAVDNQPRDIGLQIK
                      ALIKKYIQRQQTINLVVVPCNVDIATTEALSMAHEVDPEGDRTI"
     repeat_region    57922. .58215
                      /rpt_family="Alu"
     repeat_region    59599. .59877
                      /rpt_family="Alu"
     repeat_region    60428. .60810
                      /rpt_family="THE1"
     repeat_region    61246. .61366
                      /rpt_family="Alu"

  Query Match             64.0%;  Score 16;  DB 9;  Length 71230;
  Best Local Similarity   100.0%;  Pred. No. 72;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
            ||||||||||||||||
Db   40455 GTCAGGTGCAGGTCAG 40470
```

```
RESULT  32
AC002088/c
LOCUS       AC002088    85633 bp    DNA             PRI       04-FEB-2000
DEFINITION  Human BAC clone CTB-13P7 from 7p15-p21, complete sequence.
ACCESSION   AC002088
VERSION     AC002088.1  GI:2085779
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 85633)
  AUTHORS   Waterston,B. and Scheet,P.
  TITLE     The sequence of H. sapiens BAC clone CTB-13P7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 85633)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1997)
REFERENCE   3  (bases 1 to 85633)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 85633)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_RG013P07
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded or sequenced with an alternate
            chemistry; an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one subclone; and the assembly was
            confirmed by restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project and the Washington University Genome Sequencing Center.
            For additional information about the map position of this sequence,
            see http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send an E-mail to
            egreen@nhgri.nih.gov

            SOURCE INFORMATION:
            This clone is from the first release of the human BAC library.  The
            library contains cloned DNA from a human male fibroblast cell line
            978SK.  See:  Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7
            (1992); Kim et al., Genomics 34:213-8 (1996).  The clone is
            available from Research Genetics, Inc. (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of CTB-13P7;
            actual end is at 85633 of CTB-13P7.  The orientation of this clone
            is unknown.

            This clone contains sWSS2365 (NID:g1254783).
FEATURES             Location/Qualifiers
     source          1. .85633
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="CTB-13P7"
                     /clone_lib="CITB-978SK-B"
                     /map="7p15-p21"
     repeat_region   complement(3420. .3446)
                     /rpt_family="L1"
     repeat_region   complement(4016. .4306)
                     /rpt_family="ALU"
     repeat_region   complement(6902. .6942)
                     /rpt_family="L1"
     repeat_region   13108. .13401
                     /rpt_family="ALU"
     repeat_region   complement(16600. .16982)
                     /rpt_family="L1"
     repeat_region   complement(17300. .17375)
                     /rpt_family="ALU"
     repeat_region   complement(17475. .17502)
                     /rpt_family="L1"
     repeat_region   17503. .17584
                     /rpt_family="ALU"
     repeat_region   17625. .17838
                     /rpt_family="ALU"
     repeat_region   17979. .18078
                     /rpt_family="L1"
     repeat_region   complement(18407. .18454)
                     /rpt_family="L1"
     repeat_region   complement(24978. .25240)
                     /rpt_family="ALU"
     repeat_region   25293. .25377
                     /rpt_family="L1"
     repeat_region   complement(25503. .25778)
                     /rpt_family="L1"
     repeat_region   complement(25779. .26070)
                     /rpt_family="ALU"
     repeat_region   complement(26076. .26884)
                     /rpt_family="L1"
     repeat_region   complement(26887. .27177)
                     /rpt_family="ALU"
     repeat_region   complement(27242. .27327)
                     /rpt_family="L1"
     repeat_region   complement(27436. .27612)
                     /rpt_family="L1"
     repeat_region   complement(28177. .28229)
                     /rpt_family="L1"
     repeat_region   complement(28230. .28517)
                     /rpt_family="ALU"
     repeat_region   complement(28518. .28570)
                     /rpt_family="L1"
     repeat_region   31781. .32083
                     /rpt_family="MER"
     repeat_region   32102. .32130
                     /rpt_family="MER"
     repeat_region   complement(34983. .35002)
                     /rpt_family="L1"
     repeat_region   complement(35680. .35721)
                     /rpt_family="L1"
     repeat_region   complement(38336. .38360)
                     /rpt_family="L1"
     repeat_region   complement(39172. .39464)
                     /rpt_family="ALU"
     repeat_region   complement(39570. .39698)
                     /rpt_family="ALU"
     repeat_region   complement(39954. .40230)
                     /rpt_family="L1"
     repeat_region   40007. .40195
                     /rpt_family="L1"
     repeat_region   complement(40341. .40901)
```

```
                     /rpt_family="L1"
     repeat_region   40345. .40893
                     /rpt_family="L1"
     repeat_region   complement(41771. .47939)
                     /rpt_family="L1"
     repeat_region   complement(47950. .48026)
                     /rpt_family="MER"
     repeat_region   complement(48181. .48428)
                     /rpt_family="MER"
     repeat_region   48233. .48439
                     /rpt_family="MER"
     repeat_region   complement(48619. .48655)
                     /rpt_family="MER"
     repeat_region   complement(48694. .48890)
                     /rpt_family="MER"
     repeat_region   48819. .48974
                     /rpt_family="MER"
     repeat_region   49030. .49292
                     /rpt_family="MER"
     repeat_region   49440. .49757
                     /rpt_family="MER"
     repeat_region   complement(49570. .49873)
                     /rpt_family="MER"
     repeat_region   51098. .51155
                     /rpt_family="ALU"
     repeat_region   51181. .51230
                     /rpt_family="ALU"
     repeat_region   complement(51563. .51830)
                     /rpt_family="ALU"
     repeat_region   complement(52400. .52441)
                     /rpt_family="L1"
     repeat_region   complement(53064. .53085)
                     /rpt_family="L1"
     repeat_region   complement(53087. .53384)
                     /rpt_family="ALU"
     repeat_region   complement(53410. .53435)
                     /rpt_family="L1"
     repeat_region   complement(53436. .53545)
                     /rpt_family="ALU"
     repeat_region   complement(53554. .53709)
                     /rpt_family="ALU"
     repeat_region   complement(60972. .60992)
                     /rpt_family="L1"
     repeat_region   complement(64802. .64824)
                     /rpt_family="L1"
     repeat_region   complement(66753. .67258)
                     /rpt_family="MER"
     repeat_region   complement(72167. .72185)
                     /rpt_family="L1"
     repeat_region   72879. .73153
                     /rpt_family="ALU"
     repeat_region   complement(82252. .82468)
                     /rpt_family="MER"
     repeat_region   82281. .82473
                     /rpt_family="MER"
     repeat_region   82600. .82926
                     /rpt_family="MER"
     repeat_region   complement(83480. .83780)

  Query Match             64.0%;  Score 16;  DB 9;  Length 85633;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagca 20
          ||||||||||||||||
Db  22093 CAGGTGCAGGTCAGCA 22078


RESULT  33
AC007555/c
LOCUS       AC007555   109973 bp    DNA             HTG       24-AUG-2000
DEFINITION  Homo sapiens chromosome 7 clone RP5-925B6, WORKING DRAFT SEQUENCE,
            2 unordered pieces.
ACCESSION   AC007555
VERSION     AC007555.2  GI:9887766
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 109973)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 109973)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 24, 2000 this sequence version replaced gi:4827326.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_DJ0925B06
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 79%
            Sequencing vector: plasmid; 21%
            Chemistry: Dye-primer ET; 79% of reads
            Chemistry: Dye-terminator Big Dye; 21% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 109177 bases at least Q40
            Consensus quality: 109655 bases at least Q30
            Consensus quality: 109773 bases at least Q20
            Insert size: 201000; agarose-fp
            Insert size: 0; sum-of-contigs
            Quality coverage:  in Q20 bases; agarose-fp
            Quality coverage:  in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    25831: contig of 25831 bp in length
            *    25832    25931: gap of unknown length
            *    25932   109973: contig of 84042 bp in length.
FEATURES             Location/Qualifiers
     source          1. .109973
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP5-925B6"
     misc_feature    1. .25831
                     /note="assembly_name:Contig5
                     clone_end:SP6
                     vector_side:left"
     misc_feature    25932. .109973
                     /note="assembly_name:Contig6
                     clone_end:T7
                     vector_side:left"
BASE COUNT    34650 a  19714 c  19631 g  35877 t    101 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 109973;
  Best Local Similarity   100.0%;  Pred. No. 67;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagca 20
```

```
          ||||||||||||||||
Db  91300 CAGGTGCAGGTCAGCA 91285


RESULT  34
AC092845/c
LOCUS       AC092845   119749 bp    DNA             HTG       30-JUL-2001
DEFINITION  Homo sapiens chromosome UNK clone RP11-480J15, *** SEQUENCING IN
            PROGRESS ***, 46 unordered pieces.
ACCESSION   AC092845
VERSION     AC092845.1  GI:15029484
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 119749)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 119749)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------------- Project Information -----------------
            Center project name: H_NH0480J15
            ------------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 46 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1171: contig of 1171 bp in length
            *     1172     1271: gap of unknown length
            *     1272     2890: contig of 1619 bp in length
            *     2891     2990: gap of unknown length
            *     2991     4472: contig of 1482 bp in length
            *     4473     4572: gap of unknown length
            *     4573     5850: contig of 1278 bp in length
            *     5851     5950: gap of unknown length
            *     5951     7292: contig of 1342 bp in length
            *     7293     7392: gap of unknown length
            *     7393     9328: contig of 1936 bp in length
            *     9329     9428: gap of unknown length
            *     9429    11085: contig of 1657 bp in length
            *    11086    11185: gap of unknown length
            *    11186    12927: contig of 1742 bp in length
            *    12928    13027: gap of unknown length
            *    13028    14570: contig of 1543 bp in length
            *    14571    14670: gap of unknown length
            *    14671    15742: contig of 1072 bp in length
            *    15743    15842: gap of unknown length
            *    15843    17083: contig of 1241 bp in length
            *    17084    17183: gap of unknown length
            *    17184    18697: contig of 1514 bp in length
            *    18698    18797: gap of unknown length
            *    18798    19875: contig of 1078 bp in length
            *    19876    19975: gap of unknown length
            *    19976    21390: contig of 1415 bp in length
            *    21391    21490: gap of unknown length
            *    21491    23170: contig of 1680 bp in length
            *    23171    23270: gap of unknown length
            *    23271    24798: contig of 1528 bp in length
            *    24799    24898: gap of unknown length
            *    24899    27682: contig of 2784 bp in length
            *    27683    27782: gap of unknown length
            *    27783    29997: contig of 2215 bp in length
            *    29998    30097: gap of unknown length
            *    30098    32119: contig of 2022 bp in length
            *    32120    32219: gap of unknown length
            *    32220    34075: contig of 1856 bp in length
            *    34076    34175: gap of unknown length
            *    34176    36611: contig of 2436 bp in length
            *    36612    36711: gap of unknown length
            *    36712    38912: contig of 2201 bp in length
            *    38913    39012: gap of unknown length
            *    39013    40658: contig of 1646 bp in length
            *    40659    40758: gap of unknown length
            *    40759    42885: contig of 2127 bp in length
            *    42886    42985: gap of unknown length
            *    42986    44675: contig of 1690 bp in length
            *    44676    44775: gap of unknown length
            *    44776    46101: contig of 1326 bp in length
            *    46102    46201: gap of unknown length
            *    46202    47837: contig of 1636 bp in length
            *    47838    47937: gap of unknown length
            *    47938    50354: contig of 2417 bp in length
            *    50355    50454: gap of unknown length
            *    50455    53484: contig of 3030 bp in length
            *    53485    53584: gap of unknown length
            *    53585    57512: contig of 3928 bp in length
            *    57513    57612: gap of unknown length
            *    57613    59689: contig of 2077 bp in length
            *    59690    59789: gap of unknown length
            *    59790    62280: contig of 2491 bp in length
            *    62281    62380: gap of unknown length
            *    62381    65665: contig of 3285 bp in length
            *    65666    65765: gap of unknown length
            *    65766    68813: contig of 3048 bp in length
            *    68814    68913: gap of unknown length
            *    68914    72349: contig of 3436 bp in length
            *    72350    72449: gap of unknown length
            *    72450    74867: contig of 2418 bp in length
            *    74868    74967: gap of unknown length
            *    74968    77456: contig of 2489 bp in length
            *    77457    77556: gap of unknown length
            *    77557    80059: contig of 2503 bp in length
            *    80060    80159: gap of unknown length
            *    80160    82473: contig of 2314 bp in length
            *    82474    82573: gap of unknown length
            *    82574    86092: contig of 3519 bp in length
            *    86093    86192: gap of unknown length
            *    86193    89565: contig of 3373 bp in length
            *    89566    89665: gap of unknown length
            *    89666    92631: contig of 2966 bp in length
            *    92632    92731: gap of unknown length
            *    92732    96074: contig of 3343 bp in length
            *    96075    96174: gap of unknown length
            *    96175   100717: contig of 4543 bp in length
            *   100718   100817: gap of unknown length
            *   100818   107308: contig of 6491 bp in length
            *   107309   107408: gap of unknown length
            *   107409   119749: contig of 12341 bp in length.
FEATURES             Location/Qualifiers
     source          1. .119749
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="RP11-480J15"
     misc_feature    1. .1171
                     /note="assembly_name:Contig11"
     misc_feature    1272. .2890
                     /note="assembly_name:Contig27"
     misc_feature    2991. .4472
                     /note="assembly_name:Contig32"
     misc_feature    4573. .5850
```

```
                     /note="assembly_name:Contig35"
     misc_feature    5951. .7292
                     /note="assembly_name:Contig36"
     misc_feature    7393. .9328
                     /note="assembly_name:Contig37"
     misc_feature    9429. .11085
                     /note="assembly_name:Contig38"
     misc_feature    11186. .12927
                     /note="assembly_name:Contig39"
     misc_feature    13028. .14570
                     /note="assembly_name:Contig40"
     misc_feature    14671. .15742
                     /note="assembly_name:Contig41"
     misc_feature    15843. .17083
                     /note="assembly_name:Contig43"
     misc_feature    17184. .18697
                     /note="assembly_name:Contig44"
     misc_feature    18798. .19875
                     /note="assembly_name:Contig45"
     misc_feature    19976. .21390
                     /note="assembly_name:Contig47"
     misc_feature    21491. .23170
                     /note="assembly_name:Contig49"
     misc_feature    23271. .24798
                     /note="assembly_name:Contig50"
     misc_feature    24899. .27682
                     /note="assembly_name:Contig51"
     misc_feature    27783. .29997
                     /note="assembly_name:Contig52"
     misc_feature    30098. .32119
                     /note="assembly_name:Contig53"
     misc_feature    32220. .34075
                     /note="assembly_name:Contig54"
     misc_feature    34176. .36611
                     /note="assembly_name:Contig55"
     misc_feature    36712. .38912
                     /note="assembly_name:Contig56"
     misc_feature    39013. .40658
                     /note="assembly_name:Contig57"
     misc_feature    40759. .42885
                     /note="assembly_name:Contig58"
     misc_feature    42986. .44675
                     /note="assembly_name:Contig59"
     misc_feature    44776. .46101
                     /note="assembly_name:Contig60
                     clone_end:SP6
                     vector_side:left"
     misc_feature    46202. .47837
                     /note="assembly_name:Contig62"
     misc_feature    47938. .50354
                     /note="assembly_name:Contig63"
     misc_feature    50455. .53484
                     /note="assembly_name:Contig65"
     misc_feature    53585. .57512
                     /note="assembly_name:Contig66"
     misc_feature    57613. .59689
                     /note="assembly_name:Contig67"
     misc_feature    59790. .62280
                     /note="assembly_name:Contig68"
     misc_feature    62381. .65665

  Query Match             64.0%;  Score 16;  DB 2;  Length 119749;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db     930 CAGGTGCAGGTCAGCA 915


RESULT  35
AC073052/c
LOCUS       AC073052   127070 bp    DNA             PRI       09-AUG-2001
```

```
DEFINITION  Homo sapiens chromosome 2 clone RP11-92F20, complete sequence.
ACCESSION   AC073052
VERSION     AC073052.5  GI:15145540
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127070)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 127070)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 127070)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 9, 2001 this sequence version replaced gi:14249098.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH0092F20
            ------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .127070
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-92F20"
BASE COUNT    41948 a  25328 c  21891 g  37903 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 9;  Length 127070;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db 124378 CAGGTGCAGGTCAGCA 124363


RESULT  36
AC027810/c
LOCUS       AC027810   155628 bp    DNA             HTG       26-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-357J12 map 18, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC027810
VERSION     AC027810.3  GI:8082020
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155628)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-357J12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155628)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
```

```
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7684514.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9097
                Center clone name: 357_J_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 146315 bases at least Q40
                Consensus quality: 151241 bases at least Q30
                Consensus quality: 153082 bases at least Q20
                Insert size: 163000; agarose-fp
                Insert size: 154328; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 4.4 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1068: contig of 1068 bp in length
            *     1069  1168: gap of      100 bp
            *     1169     2717: contig of 1549 bp in length
            *     2718  2817: gap of      100 bp
            *     2818     6040: contig of 3223 bp in length
            *     6041  6140: gap of      100 bp
            *     6141    10201: contig of 4061 bp in length
            *    10202 10301: gap of      100 bp
            *    10302    14615: contig of 4314 bp in length
            *    14616 14715: gap of      100 bp
            *    14716    21086: contig of 6371 bp in length
            *    21087 21186: gap of      100 bp
            *    21187    25767: contig of 4581 bp in length
            *    25768 25867: gap of      100 bp
            *    25868    32500: contig of 6633 bp in length
            *    32501 32600: gap of      100 bp
            *    32601    38651: contig of 6051 bp in length
            *    38652 38751: gap of      100 bp
            *    38752    50922: contig of 12171 bp in length
            *    50923 51022: gap of      100 bp
            *    51023    71014: contig of 19992 bp in length
            *    71015 71114: gap of      100 bp
            *    71115    93563: contig of 22449 bp in length
            *    93564 93663: gap of      100 bp
            *    93664   123080: contig of 29417 bp in length
            *   123081 123180: gap of      100 bp
            *   123181   155628: contig of 32448 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155628
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-357J12"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1068
                     /note="assembly_fragment"
     misc_feature    1169. .2717
                     /note="assembly_fragment"
     misc_feature    2818. .6040
                     /note="assembly_fragment"
     misc_feature    6141. .10201
                     /note="assembly_fragment"
     misc_feature    10302. .14615
                     /note="assembly_fragment"
     misc_feature    14716. .21086
                     /note="assembly_fragment"
     misc_feature    21187. .25767
                     /note="assembly_fragment"
     misc_feature    25868. .32500
                     /note="assembly_fragment"
     misc_feature    32601. .38651
                     /note="assembly_fragment"
     misc_feature    38752. .50922
                     /note="assembly_fragment"
     misc_feature    51023. .71014
                     /note="assembly_fragment"
     misc_feature    71115. .93563
                     /note="assembly_fragment"
     misc_feature    93664. .123080
                     /note="assembly_fragment"
     misc_feature    123181. .155628
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
BASE COUNT    37665 a  38986 c  40608 g  37067 t   1302 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 155628;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db   95800 GTCAGGTGCAGGTCAG 95785


RESULT  37
AC073247
LOCUS       AC073247   156119 bp    DNA             HTG       12-AUG-2000
DEFINITION  Homo sapiens chromosome 15 clone CTD-2006N8, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AC073247
VERSION     AC073247.2  GI:8571764
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156119)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156119)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 18, 2000 this sequence version replaced gi:8468936.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_MS2006N08
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 139572 bases at least Q40
            Consensus quality: 144812 bases at least Q30
            Consensus quality: 147446 bases at least Q20
            Insert size: 160000; agarose-fp
            Insert size: 154019; sum-of-contigs
            Quality coverage: 4.17 in Q20 bases; agarose-fp
            Quality coverage: 4.44 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1286: contig of 1286 bp in length
            *     1287     1386: gap of unknown length
            *     1387     3603: contig of 2217 bp in length
            *     3604     3703: gap of unknown length
            *     3704     6043: contig of 2340 bp in length
            *     6044     6143: gap of unknown length
            *     6144     9354: contig of 3211 bp in length
            *     9355     9454: gap of unknown length
            *     9455    12077: contig of 2623 bp in length
            *    12078    12177: gap of unknown length
            *    12178    15985: contig of 3808 bp in length
            *    15986    16085: gap of unknown length
            *    16086    19801: contig of 3716 bp in length
            *    19802    19901: gap of unknown length
            *    19902    25990: contig of 6089 bp in length
            *    25991    26090: gap of unknown length
            *    26091    30598: contig of 4508 bp in length
            *    30599    30698: gap of unknown length
            *    30699    36263: contig of 5565 bp in length
            *    36264    36363: gap of unknown length
            *    36364    42783: contig of 6420 bp in length
            *    42784    42883: gap of unknown length
            *    42884    49478: contig of 6595 bp in length
            *    49479    49578: gap of unknown length
            *    49579    56751: contig of 7173 bp in length
            *    56752    56851: gap of unknown length
            *    56852    61885: contig of 5034 bp in length
            *    61886    61985: gap of unknown length
            *    61986    69384: contig of 7399 bp in length
            *    69385    69484: gap of unknown length
            *    69485    76911: contig of 7427 bp in length
            *    76912    77011: gap of unknown length
            *    77012    86274: contig of 9263 bp in length
            *    86275    86374: gap of unknown length
            *    86375    96575: contig of 10201 bp in length
            *    96576    96675: gap of unknown length
            *    96676   109456: contig of 12781 bp in length
            *   109457   109556: gap of unknown length
            *   109557   123883: contig of 14327 bp in length
            *   123884   123983: gap of unknown length
            *   123984   138107: contig of 14124 bp in length
            *   138108   138207: gap of unknown length
            *   138208   156119: contig of 17912 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156119
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="CTD-2006N8"
BASE COUNT    42637 a  33200 c  33979 g  44179 t   2124 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 156119;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   90601 CAGGTGCAGGTCAGCA 90616


RESULT  38
AC040930
LOCUS       AC040930   162043 bp    DNA              HTG       11-JUN-2001
DEFINITION  Homo sapiens chromosome 15 clone CTD-2006N8 map 15, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC040930
VERSION     AC040930.3  GI:12658160
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162043)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone CTD-2006N8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162043)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 3, 2001 this sequence version replaced gi:7705184.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
```

```
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
            Center project name: L8890
            Center clone name: 2006_N_8
          -------------- Summary Statistics
            Sequencing vector: M13; M77815; 47% of reads
            Sequencing vector: Plasmid; n/a; 53% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 158666 bases at least Q40
            Consensus quality: 159912 bases at least Q30
            Consensus quality: 160439 bases at least Q20
            Insert size: 157000; agarose-fp
            Insert size: 161043; sum-of-contigs
            Quality coverage: 10.5 in Q20 bases; agarose-fp
            Quality coverage: 10.2 in Q20.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 11 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1      840: contig of 840 bp in length
          *      841 940: gap of      100 bp
          *      941     1652: contig of 712 bp in length
          *     1653 1752: gap of      100 bp
          *     1753     2845: contig of 1093 bp in length
          *     2846 2945: gap of      100 bp
          *     2946     4039: contig of 1094 bp in length
          *     4040 4139: gap of      100 bp
          *     4140     7025: contig of 2886 bp in length
          *     7026 7125: gap of      100 bp
          *     7126    10842: contig of 3717 bp in length
          *    10843 10942: gap of      100 bp
          *    10943    18373: contig of 7431 bp in length
          *    18374 18473: gap of      100 bp
          *    18474    28208: contig of 9735 bp in length
          *    28209 28308: gap of      100 bp
          *    28309    48596: contig of 20288 bp in length
          *    48597 48696: gap of      100 bp
          *    48697    76612: contig of 27916 bp in length
          *    76613 76712: gap of      100 bp
          *    76713   162043: contig of 85331 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162043
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="CTD-2006N8"
                     /clone_lib="CITD Human BAC"
     misc_feature    1. .840
                     /note="assembly_fragment"
     misc_feature    941. .1652
                     /note="assembly_fragment"
     misc_feature    1753. .2845
                     /note="assembly_fragment"
     misc_feature    2946. .4039
                     /note="assembly_fragment"
     misc_feature    4140. .7025
                     /note="assembly_fragment"
     misc_feature    7126. .10842
                     /note="assembly_fragment"
     misc_feature    10943. .18373
                     /note="assembly_fragment"
     misc_feature    18474. .28208
                     /note="assembly_fragment"
     misc_feature    28309. .48596
                     /note="assembly_fragment"
     misc_feature    48697. .76612
                     /note="assembly_fragment"
     misc_feature    76713. .162043
                     /note="assembly_fragment"
BASE COUNT    44171 a  35675 c  34546 g  46650 t   1001 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 162043;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   82976 CAGGTGCAGGTCAGCA 82991


RESULT  39
AC068538/c
LOCUS       AC068538   175459 bp    DNA             HTG       13-AUG-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-141G4, WORKING DRAFT SEQUENCE,
            29 unordered pieces.
ACCESSION   AC068538
VERSION     AC068538.5  GI:9795677
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175459)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175459)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 11, 2000 this sequence version replaced gi:9690368.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0141G04
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 158616 bases at least Q40
            Consensus quality: 162423 bases at least Q30
            Consensus quality: 164318 bases at least Q20
            Insert size: 179000; agarose-fp
            Insert size: 172659; sum-of-contigs
            Quality coverage: 4.51 in Q20 bases; agarose-fp
            Quality coverage: 4.87 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1355: contig of 1355 bp in length
            *     1356     1455: gap of unknown length
            *     1456     3082: contig of 1627 bp in length
            *     3083     3182: gap of unknown length
```

```
            *      3183      4263: contig of 1081 bp in length
            *      4264      4363: gap of unknown length
            *      4364      5626: contig of 1263 bp in length
            *      5627      5726: gap of unknown length
            *      5727      7533: contig of 1807 bp in length
            *      7534      7633: gap of unknown length
            *      7634      9202: contig of 1569 bp in length
            *      9203      9302: gap of unknown length
            *      9303     11061: contig of 1759 bp in length
            *     11062     11161: gap of unknown length
            *     11162     14017: contig of 2856 bp in length
            *     14018     14117: gap of unknown length
            *     14118     16928: contig of 2811 bp in length
            *     16929     17028: gap of unknown length
            *     17029     19922: contig of 2894 bp in length
            *     19923     20022: gap of unknown length
            *     20023     23986: contig of 3964 bp in length
            *     23987     24086: gap of unknown length
            *     24087     27864: contig of 3778 bp in length
            *     27865     27964: gap of unknown length
            *     27965     31695: contig of 3731 bp in length
            *     31696     31795: gap of unknown length
            *     31796     34671: contig of 2876 bp in length
            *     34672     34771: gap of unknown length
            *     34772     39459: contig of 4688 bp in length
            *     39460     39559: gap of unknown length
            *     39560     43565: contig of 4006 bp in length
            *     43566     43665: gap of unknown length
            *     43666     47318: contig of 3653 bp in length
            *     47319     47418: gap of unknown length
            *     47419     53986: contig of 6568 bp in length
            *     53987     54086: gap of unknown length
            *     54087     61186: contig of 7100 bp in length
            *     61187     61286: gap of unknown length
            *     61287     66331: contig of 5045 bp in length
            *     66332     66431: gap of unknown length
            *     66432     72447: contig of 6016 bp in length
            *     72448     72547: gap of unknown length
            *     72548     80701: contig of 8154 bp in length
            *     80702     80801: gap of unknown length
            *     80802     89408: contig of 8607 bp in length
            *     89409     89508: gap of unknown length
            *     89509    100140: contig of 10632 bp in length
            *    100141    100240: gap of unknown length
            *    100241    109860: contig of 9620 bp in length
            *    109861    109960: gap of unknown length
            *    109961    120505: contig of 10545 bp in length
            *    120506    120605: gap of unknown length
            *    120606    135065: contig of 14460 bp in length
            *    135066    135165: gap of unknown length
            *    135166    153777: contig of 18612 bp in length
            *    153778    153877: gap of unknown length
            *    153878    175459: contig of 21582 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175459
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-141G4"
     misc_feature    1. .1355
                     /note="assembly_name:Contig9"
     misc_feature    1456. .3082
                     /note="assembly_name:Contig11"
     misc_feature    3183. .4263
                     /note="assembly_name:Contig13"
     misc_feature    4364. .5626
                     /note="assembly_name:Contig14"
     misc_feature    5727. .7533
                     /note="assembly_name:Contig15"
     misc_feature    7634. .9202
                     /note="assembly_name:Contig16
                     clone_end:T7
                     vector_side:left"
     misc_feature    9303. .11061
                     /note="assembly_name:Contig17"
     misc_feature    11162. .14017
                     /note="assembly_name:Contig18"
     misc_feature    14118. .16928
                     /note="assembly_name:Contig19"
     misc_feature    17029. .19922
                     /note="assembly_name:Contig20"
     misc_feature    20023. .23986
                     /note="assembly_name:Contig21"
     misc_feature    24087. .27864
                     /note="assembly_name:Contig22"
     misc_feature    27965. .31695
                     /note="assembly_name:Contig23"
     misc_feature    31796. .34671
                     /note="assembly_name:Contig24"
     misc_feature    34772. .39459
                     /note="assembly_name:Contig25"
     misc_feature    39560. .43565
                     /note="assembly_name:Contig26"
     misc_feature    43666. .47318
                     /note="assembly_name:Contig27"
     misc_feature    47419. .53986
                     /note="assembly_name:Contig28"
     misc_feature    54087. .61186
                     /note="assembly_name:Contig29"
     misc_feature    61287. .66331
                     /note="assembly_name:Contig30"
     misc_feature    66432. .72447
                     /note="assembly_name:Contig31
                     clone_end:SP6
                     vector_side:right"
     misc_feature    72548. .80701
                     /note="assembly_name:Contig32"
     misc_feature    80802. .89408
                     /note="assembly_name:Contig33"
     misc_feature    89509. .100140
                     /note="assembly_name:Contig34"
     misc_feature    100241. .109860
                     /note="assembly_name:Contig35"
     misc_feature    109961. .120505
                     /note="assembly_name:Contig36"
     misc_feature    120606. .135065
                     /note="assembly_name:Contig37"
     misc_feature    135166. .153777
                     /note="assembly_name:Contig38"
     misc_feature    153878. .175459
                     /note="assembly_name:Contig39"
BASE COUNT    47828 a  38505 c  38363 g  47961 t   2802 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 175459;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   79767 CAGGTGCAGGTCAGCA 79752


RESULT  40
AL139326/c
LOCUS       AL139326   175588 bp    DNA             PRI       09-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-351K3 on chromosome 13, complete
            sequence.
ACCESSION   AL139326
VERSION     AL139326.15  GI:14018249
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 175588)
  AUTHORS   Blakey,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 11, 2001 this sequence version replaced gi:13277016.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-351K3 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-351K3 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-351K3 is at 1 in this sequence. The
            true left end of clone RP11-154C3 is at 175489 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .175588
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-351K3"
                     /clone_lib="RPCI-11.2"
     repeat_region   68. .360
                     /note="AluSx repeat: matches 1. .290 of consensus"
     repeat_region   953. .1089
                     /note="MIR repeat: matches 80. .210 of consensus"
     repeat_region   3306. .3437
                     /note="L1MD1 repeat: matches 6066. .6201 of consensus"
     repeat_region   3591. .3719
                     /note="MER47 repeat: matches 1. .193 of consensus"
     repeat_region   3722. .3797
                     /note="MER47 repeat: matches 2248. .2323 of consensus"
     repeat_region   3803. .3878
                     /note="L1MEc repeat: matches 2203. .2279 of consensus"
     repeat_region   3879. .4182
                     /note="AluJo repeat: matches 1. .289 of consensus"
     repeat_region   4183. .4479
                     /note="L1MEc repeat: matches 1903. .2203 of consensus"
     repeat_region   4480. .4788
                     /note="AluSq repeat: matches 2. .310 of consensus"
     repeat_region   4789. .5617
                     /note="L1MEc repeat: matches 1094. .1903 of consensus"
     repeat_region   5628. .5925
                     /note="AluY repeat: matches 1. .298 of consensus"
     repeat_region   6403. .6459
                     /note="MIR repeat: matches 197. .259 of consensus"
     repeat_region   6899. .7218
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   7225. .7364
                     /note="MIR repeat: matches 46. .188 of consensus"
     repeat_region   7961. .8260
                     /note="AluJb repeat: matches 1. .300 of consensus"
     repeat_region   8440. .8736
                     /note="AluJb repeat: matches 1. .298 of consensus"
     repeat_region   9747. .9956
                     /note="MIR repeat: matches 36. .250 of consensus"
     repeat_region   10124. .10315
                     /note="L1ME1 repeat: matches 5988. .6322 of consensus"
     repeat_region   10384. .10682
                     /note="AluY repeat: matches 1. .300 of consensus"
     repeat_region   10954. .11235
                     /note="AluY repeat: matches 1. .294 of consensus"
     repeat_region   11236. .11543
                     /note="AluSx repeat: matches 3. .294 of consensus"
     repeat_region   12286. .12493
                     /note="MIR repeat: matches 32. .249 of consensus"
     repeat_region   13328. .13622
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   15766. .15864
                     /note="L2 repeat: matches 2630. .2736 of consensus"
     repeat_region   16580. .16778
                     /note="L2 repeat: matches 2540. .2742 of consensus"
     repeat_region   17117. .17212
                     /note="MIR repeat: matches 30. .124 of consensus"
     repeat_region   17358. .18333
                     /note="SVA repeat: matches 2. .954 of consensus"
     misc_feature    17886. .19097
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    18166. .18189
                     /note="1372 bases of IS186 (X03123) removed here.  This
                     sequence represents the duplicated flanking sequence of
                     the IS186."
     repeat_region   18343. .19109
                     /note="SVA repeat: matches 646. .1386 of consensus"
     repeat_region   19813. .21277
                     /note="L1MA4 repeat: matches 4785. .6297 of consensus"
     repeat_region   21663. .21969
                     /note="AluJb repeat: matches 1. .308 of consensus"
     repeat_region   22425. .22651
                     /note="L1M4 repeat: matches 4456. .4697 of consensus"
     repeat_region   23061. .23181
                     /note="MIR repeat: matches 34. .170 of consensus"
     repeat_region   23287. .23332
                     /note="L2 repeat: matches 2692. .2737 of consensus"
     repeat_region   23293. .23345
                     /note="MIR repeat: matches 200. .249 of consensus"
     repeat_region   26800. .27101
                     /note="AluSq repeat: matches 1. .302 of consensus"
     repeat_region   27275. .27582
                     /note="AluJb repeat: matches 1. .306 of consensus"
     repeat_region   27607. .27900
                     /note="AluSc repeat: matches 1. .308 of consensus"
     repeat_region   27909. .28044
                     /note="FLAM_C repeat: matches 1. .133 of consensus"
     repeat_region   28130. .28431
                     /note="AluJb repeat: matches 6. .305 of consensus"
     repeat_region   28517. .28572
                     /note="MIR repeat: matches 96. .155 of consensus"
     repeat_region   29159. .29275
                     /note="AluJb repeat: matches 1. .118 of consensus"
     repeat_region   29425. .29729
                     /note="AluY repeat: matches 1. .305 of consensus"
     repeat_region   30098. .30173
                     /note="L1MC5 repeat: matches 7840. .7913 of consensus"
     repeat_region   30259. .30544
                     /note="AluJb repeat: matches 1. .284 of consensus"
     repeat_region   30739. .30875
                     /note="L1MC5 repeat: matches 7603. .7731 of consensus"
     repeat_region   30860. .31103
                     /note="L1MC4 repeat: matches 7643. .7901 of consensus"
     repeat_region   31112. .31223
                     /note="AluSp/q repeat: matches 194. .305 of consensus"
```

```
     repeat_region   31252. .31598
                     /note="L1M4 repeat: matches 4649. .4974 of consensus"
     repeat_region   31646. .32054
                     /note="L1MC4 repeat: matches 6419. .6828 of consensus"
     repeat_region   32497. .32783
                     /note="AluJb repeat: matches 1. .291 of consensus"
     repeat_region   32784. .32851
                     /note="34 copies 2 mer tg 89% conserved"
     repeat_region   32852. .32873
                     /note="11 copies 2 mer ta 100% conserved"
     repeat_region   33312. .34465
                     /note="L1MC3 repeat: matches 6641. .7739 of consensus"
     repeat_region   34501. .34531
                     /note="L1MC/D repeat: matches 5494. .5524 of consensus"
     repeat_region   35395. .35581
                     /note="L1PB2 repeat: matches 5958. .6152 of consensus"
     repeat_region   37300. .37451
                     /note="L2 repeat: matches 1929. .2081 of consensus"
     repeat_region   37499. .37661
                     /note="MER33 repeat: matches -8. .156 of consensus"
     repeat_region   37662. .37968
                     /note="AluSx repeat: matches 1. .307 of consensus"
     repeat_region   37969. .38061
                     /note="MER33 repeat: matches 156. .243 of consensus"
     repeat_region   38063. .38361
                     /note="AluSx repeat: matches 1. .296 of consensus"
     repeat_region   38434. .38651
                     /note="AluJo repeat: matches 1. .217 of consensus"
     repeat_region   38654. .38833
                     /note="AluSg/x repeat: matches 121. .299 of consensus"
     repeat_region   38947. .39134
                     /note="L2 repeat: matches 1470. .1655 of consensus"
     repeat_region   40240. .40544
                     /note="AluSx repeat: matches 1. .305 of consensus"
     repeat_region   40547. .40855
                     /note="AluSg repeat: matches 3. .310 of consensus"
     repeat_region   41027. .41160
                     /note="L1MC4 repeat: matches 7837. .7977 of consensus"
     repeat_region   41161. .41198
                     /note="19 copies 2 mer ac 86% conserved"
     repeat_region   41204. .41241
                     /note="19 copies 2 mer tg 81% conserved"
     repeat_region   41252. .41356
                     /note="L1MC4 repeat: matches 7718. .7832 of consensus"

  Query Match             64.0%;  Score 16;  DB 9;  Length 175588;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagc 19
          ||||||||||||||||
Db   9769 TCAGGTGCAGGTCAGC 9754


RESULT  41
AC092679/c
LOCUS       AC092679   182396 bp    DNA             HTG       19-JUL-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-676D22, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC092679 AC025339
VERSION     AC092679.1  GI:14916263
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182396)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182396)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 19, 2001 this sequence version replaced gi:7210047.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------------- Project Information -----------------
            Center project name: H_NH0676D22
            Drafting center: WIBR
            ------------------------------------------------------
            -------------- Summary Statistics  -------------
            Sequencing vector: M13; 52%
            Sequencing vector: plasmid; 44%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 96% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 180399 bases at least Q40
            Consensus quality: 181020 bases at least Q30
            Consensus quality: 181438 bases at least Q20
            Insert size: 178000; agarose-fp
            Insert size: 182296; sum-of-contigs
            Quality coverage: 7.87 in Q20 bases; agarose-fp
            Quality coverage: 7.71 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    82729: contig of 82729 bp in length
            *    82730    82829: gap of unknown length
            *    82830   182396: contig of 99567 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182396
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-676D22"
     misc_feature    1. .82729
                     /note="assembly_name:Contig33
                     clone_end:T7
                     vector_side:left"
     misc_feature    82830. .182396
                     /note="assembly_name:Contig34
                     clone_end:SP6
                     vector_side:left"
BASE COUNT    57381 a  38224 c  34240 g  52447 t    104 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 182396;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagca 20
          ||||||||||||||||
Db 173146 CAGGTGCAGGTCAGCA 173131


RESULT  42
AC005412/c
LOCUS       AC005412   183700 bp    DNA             PRI       22-NOV-2000
DEFINITION  Homo sapiens chromosome 17, clone hRPK.22_N_12, complete sequence.
ACCESSION   AC005412
VERSION     AC005412.6  GI:11128436
KEYWORDS    HTG.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183700)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.22_N_12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183700)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W.,
            Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraigery,K., Grant,G.,
            Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C.,
            Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczky,J.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
            Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 183700)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 183700)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 9, 2000 this sequence version replaced gi:5931398.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .183700
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="hRPK.22_N_12"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1. .70
                     /rpt_family="MER41D"
     repeat_region   784. .1084
                     /rpt_family="AluY"
     repeat_region   1320. .1637
                     /rpt_family="AluJo"
     repeat_region   complement(1664. .1959)
                     /rpt_family="AluSx"
     repeat_region   1961. .2074
                     /rpt_family="L2"
     repeat_region   2283. .2355
                     /rpt_family="C-rich"
     repeat_region   complement(2369. .2774)
                     /rpt_family="L1M4"
     repeat_region   2775. .3066
                     /rpt_family="AluSq"
     repeat_region   complement(3067. .3077)
                     /rpt_family="L1M4"
     repeat_region   3549. .3814
                     /rpt_family="AluSq"
     repeat_region   complement(3867. .3949)
                     /rpt_family="AluJo"
     repeat_region   complement(3950. .4247)
                     /rpt_family="AluJb"
     repeat_region   complement(4248. .4361)
                     /rpt_family="AluJo"
     repeat_region   complement(4362. .4490)
                     /rpt_family="L1M4"
     repeat_region   complement(4493. .4589)
                     /rpt_family="L1M4"
     repeat_region   4590. .4888
                     /rpt_family="AluSx"
     repeat_region   complement(4889. .5305)
                     /rpt_family="L1M4"
     repeat_region   complement(5328. .5623)
                     /rpt_family="AluSx"
     repeat_region   complement(5876. .6164)
                     /rpt_family="AluSg"
     repeat_region   6196. .6509
                     /rpt_family="AluSg"
     repeat_region   6525. .6580
                     /rpt_family="AT_rich"
     repeat_region   complement(6671. .6784)
                     /rpt_family="L2"
     repeat_region   complement(6986. .7137)
                     /rpt_family="AluSg"
     repeat_region   complement(7138. .7447)
                     /rpt_family="AluSx"
     repeat_region   complement(7448. .7582)
                     /rpt_family="AluSg"
     repeat_region   7808. .8106
                     /rpt_family="AluJb"
     unsure          7825. .8084
                     /note="Single-stranded terminator coverage."
     repeat_region   8121. .8476
                     /rpt_family="L1MB5"
     repeat_region   8496. .8677
                     /rpt_family="(TATATG)n"
     unsure          8503. .8791
                     /note="Single-stranded terminator coverage."
     repeat_region   8689. .8870
                     /rpt_family="(TA)n"
     unsure          8736
                     /note="Probably T, possibly C."
     repeat_region   8877. .9144
                     /rpt_family="(TATATG)n"
     repeat_region   9147. .9210
                     /rpt_family="(TA)n"
     repeat_region   9216. .9395
                     /rpt_family="(CATATA)n"
     unsure          complement(9240. .9299)
```

```
                     /note="Single-stranded terminator coverage."
     repeat_region   9458. .9530
                     /rpt_family="L1MB5"
     repeat_region   9536. .9848
                     /rpt_family="AluSx"
     repeat_region   9878. .10115
                     /rpt_family="L2"
     repeat_region   10128. .10357
                     /rpt_family="MER20"
     repeat_region   10432. .10458
                     /rpt_family="AT_rich"
     repeat_region   complement(10511. .10567)
                     /rpt_family="AluJ/FLAM"
     repeat_region   10612. .10908
                     /rpt_family="AluJb"
     repeat_region   10913. .11210
                     /rpt_family="AluSp"
     repeat_region   11216. .11512
                     /rpt_family="AluSx"
     repeat_region   11529. .11553
                     /rpt_family="(TGAA)n"
     repeat_region   complement(11647. .11759)
                     /rpt_family="L2"
     repeat_region   complement(11816. .12159)
                     /rpt_family="L2"
     repeat_region   complement(13008. .13306)
                     /rpt_family="AluSp"
     repeat_region   13355. .13490
                     /rpt_family="AluSx"
     repeat_region   13491. .13789
                     /rpt_family="AluSc"
     repeat_region   13790. .13956
                     /rpt_family="AluSx"
     repeat_region   complement(14078. .14206)
                     /rpt_family="MIR"
     repeat_region   complement(14220. .14276)
                     /rpt_family="L2"
     repeat_region   complement(14277. .14309)
                     /rpt_family="MIR"
     repeat_region   complement(14310. .14594)
                     /rpt_family="AluSp"
     repeat_region   complement(14604. .14912)
                     /rpt_family="AluJo"
     repeat_region   complement(14913. .15045)
                     /rpt_family="MIR"
     repeat_region   complement(15061. .15157)
                     /rpt_family="MIR"
     repeat_region   complement(15517. .15637)

  Query Match             64.0%;  Score 16;  DB 9;  Length 183700;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db 177274 GTCAGGTGCAGGTCAG 177259


RESULT  43
AC025339/c
ID   AC025339   standard; DNA; HTG; 185492 BP.
XX
AC   AC025339;
XX
SV   AC025339.1
XX
DT   14-MAR-2000 (Rel. 63, Created)
DT   14-MAR-2000 (Rel. 63, Last updated, Version 1)
XX
DE   Homo sapiens chromosome 2 clone RP11-676D22 map 2, WORKING DRAFT SEQUENCE,
DE   51 unordered pieces.
XX
KW   HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-185492
RA   Birren B., Linton L., Nusbaum C., Lander E.;
RT   "Homo sapiens chromosome 2, clone RP11-676D22";
RL   Unpublished.
XX
RN   [2]
RP   1-185492
RA   Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
RA   Anderson S., Baldwin J., Barna N., Bastien V., Beda F., Boguslavkiy L.,
RA   Boukhgalter B., Brown A., Burkett G., Campopiano A., Castle A., Choepel Y.,
RA   Colangelo M., Collins S., Collymore A., Cooke P., DeArellano K., Dewar K.,
RA   Diaz J.S., Dodge S., Domino M., Doyle M., Ferreira P., FitzHugh W.,
RA   Gage D., Galagan J., Gardyna S., Ginde S., Goyette M., Graham L.,
RA   Grand-Pierre N., Grant G., Hagos B., Heaford A., Horton L., Howland J.C.,
RA   Iliev I., Johnson R., Jones C., Kann L., Karatas A., Klein J., LaRocque K.,
RA   Lamazares R., Landers T., Lehoczky J., Levine R., Lieu C., Liu G.,
RA   Locke K., Macdonald P., Marquis N., McCarthy M., McEwan P., McGurk A.,
RA   McKernan K., McPheeters R., Meldrim J., Meneus L., Mihova T., Miranda C.,
RA   Mlenga V., Morrow J., Murphy T., Naylor J., Norman C.H., O'Connor T.,
RA   O'Donnell P., O'Neil D., Olivar T.M., Oliver J., Peterson K., Pierre N.,
RA   Pisani C., Pollara V., Raymond C., Riley R., Rogov P., Rothman D., Roy A.,
RA   Santos R., Schauer S., Severy P., Spencer B., Stange-Thomann N.,
RA   Stojanovic N., Subramanian A., Talamas J., Tesfaye S., Theodore J.,
RA   Tirrell A., Travers M., Trigilio J., Vassiliev H., Viel R., Vo A.,
RA   Wilson B., Wu X., Wyman D., Ye W.J., Young G., Zainoun J., Zimmer A.,
RA   Zody M.;
RT   ;
RL   Submitted (08-MAR-2000) to the EMBL/GenBank/DDBJ databases.
RL   Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL   Cambridge, MA 02141, USA
XX
CC   All repeats were identified using RepeatMasker:
CC   Smit, A.F.A. & Green, P. (1996-1997)
CC   http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC   -------------- Genome Center
CC   Center: Whitehead Institute/ MIT Center for Genome Research
CC   Center code: WIBR
CC   Web site: http://www-seq.wi.mit.edu
CC   Contact: sequence_submissions@genome.wi.mit.edu
CC   -------------- Project Information
CC   Center project name: L6951
CC   Center clone name: 676_D_22
CC   -------------- Summary Statistics
CC   Sequencing vector: M13; M77815; 100% of reads
CC   Chemistry: Dye-terminator Big Dye; 100% of reads
CC   Assembly program: Phrap; version 0.960731
CC   Consensus quality: 144276 bases at least Q40
CC   Consensus quality: 161970 bases at least Q30
CC   Consensus quality: 171080 bases at least Q20
CC   Insert size: 194000; agarose-fp
CC   Insert size: 180492; sum-of-contigs
CC   Quality coverage: 2.9 in Q20 bases; agarose-fp
CC   Quality coverage: 3.1 in Q20 bases; sum-of-contigs
CC   --------------.
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 51 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1     1026: contig of 1026 bp in length
CC   *     1027 1126: gap of      100 bp
CC   *     1127     2310: contig of 1184 bp in length
CC   *     2311 2410: gap of      100 bp
CC   *     2411     3533: contig of 1123 bp in length
```

```
CC   *     3534 3633: gap of      100 bp
CC   *     3634    4974: contig of 1341 bp in length
CC   *     4975 5074: gap of      100 bp
CC   *     5075    6187: contig of 1113 bp in length
CC   *     6188 6287: gap of      100 bp
CC   *     6288    7383: contig of 1096 bp in length
CC   *     7384 7483: gap of      100 bp
CC   *     7484    8508: contig of 1025 bp in length
CC   *     8509 8608: gap of      100 bp
CC   *     8609    9799: contig of 1191 bp in length
CC   *     9800 9899: gap of      100 bp
CC   *     9900   11275: contig of 1376 bp in length
CC   *    11276 11375: gap of      100 bp
CC   *    11376   12804: contig of 1429 bp in length
CC   *    12805 12904: gap of      100 bp
CC   *    12905   14246: contig of 1342 bp in length
CC   *    14247 14346: gap of      100 bp
CC   *    14347   15544: contig of 1198 bp in length
CC   *    15545 15644: gap of      100 bp
CC   *    15645   17181: contig of 1537 bp in length
CC   *    17182 17281: gap of      100 bp
CC   *    17282   19392: contig of 2111 bp in length
CC   *    19393 19492: gap of      100 bp
CC   *    19493   21710: contig of 2218 bp in length
CC   *    21711 21810: gap of      100 bp
CC   *    21811   24149: contig of 2339 bp in length
CC   *    24150 24249: gap of      100 bp
CC   *    24250   26685: contig of 2436 bp in length
CC   *    26686 26785: gap of      100 bp
CC   *    26786   28327: contig of 1542 bp in length
CC   *    28328 28427: gap of      100 bp
CC   *    28428   31535: contig of 3108 bp in length
CC   *    31536 31635: gap of      100 bp
CC   *    31636   34165: contig of 2530 bp in length
CC   *    34166 34265: gap of      100 bp
CC   *    34266   37131: contig of 2866 bp in length
CC   *    37132 37231: gap of      100 bp
CC   *    37232   40119: contig of 2888 bp in length
CC   *    40120 40219: gap of      100 bp
CC   *    40220   43234: contig of 3015 bp in length
CC   *    43235 43334: gap of      100 bp
CC   *    43335   46125: contig of 2791 bp in length
CC   *    46126 46225: gap of      100 bp
CC   *    46226   48748: contig of 2523 bp in length
CC   *    48749 48848: gap of      100 bp
CC   *    48849   53250: contig of 4402 bp in length
CC   *    53251 53350: gap of      100 bp
CC   *    53351   57172: contig of 3822 bp in length
CC   *    57173 57272: gap of      100 bp
CC   *    57273   60281: contig of 3009 bp in length
CC   *    60282 60381: gap of      100 bp
CC   *    60382   63484: contig of 3103 bp in length
CC   *    63485 63584: gap of      100 bp
CC   *    63585   67025: contig of 3441 bp in length
CC   *    67026 67125: gap of      100 bp
CC   *    67126   70475: contig of 3350 bp in length
CC   *    70476 70575: gap of      100 bp
CC   *    70576   74638: contig of 4063 bp in length
CC   *    74639 74738: gap of      100 bp
CC   *    74739   78257: contig of 3519 bp in length
CC   *    78258 78357: gap of      100 bp
CC   *    78358   83359: contig of 5002 bp in length
CC   *    83360 83459: gap of      100 bp
CC   *    83460   86986: contig of 3527 bp in length
CC   *    86987 87086: gap of      100 bp
CC   *    87087   90187: contig of 3101 bp in length
CC   *    90188 90287: gap of      100 bp
CC   *    90288   94735: contig of 4448 bp in length
CC   *    94736 94835: gap of      100 bp
CC   *    94836   99768: contig of 4933 bp in length
CC   *    99769 99868: gap of      100 bp
CC   *    99869  104278: contig of 4410 bp in length
CC   *   104279 104378: gap of      100 bp
CC   *   104379  110808: contig of 6430 bp in length
CC   *   110809 110908: gap of      100 bp
CC   *   110909  116478: contig of 5570 bp in length
CC   *   116479 116578: gap of      100 bp
CC   *   116579  122443: contig of 5865 bp in length
CC   *   122444 122543: gap of      100 bp
CC   *   122544  126331: contig of 3788 bp in length
CC   *   126332 126431: gap of      100 bp
CC   *   126432  131374: contig of 4943 bp in length
CC   *   131375 131474: gap of      100 bp
CC   *   131475  138177: contig of 6703 bp in length
CC   *   138178 138277: gap of      100 bp
CC   *   138278  144292: contig of 6015 bp in length
CC   *   144293 144392: gap of      100 bp
CC   *   144393  150325: contig of 5933 bp in length
CC   *   150326 150425: gap of      100 bp
CC   *   150426  156168: contig of 5743 bp in length
CC   *   156169 156268: gap of      100 bp
CC   *   156269  163498: contig of 7230 bp in length
CC   *   163499 163598: gap of      100 bp
CC   *   163599  173145: contig of 9547 bp in length
CC   *   173146 173245: gap of      100 bp
CC   *   173246  185492: contig of 12247 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .185492
FT                   /chromosome="2"
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /map="2"
FT                   /clone="RP11-676D22"
FT                   /clone_lib="RPCI-11 Human Male BAC"
FT   misc_feature    1. .1026
FT                   /note="assembly_fragment"
FT   misc_feature    1127. .2310
FT                   /note="assembly_fragment"
FT   misc_feature    2411. .3533
FT                   /note="assembly_fragment"
FT   misc_feature    3634. .4974
FT                   /note="assembly_fragment"

  Query Match             64.0%;  Score 16;  DB 33;  Length 185492;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   43512 CAGGTGCAGGTCAGCA 43497


RESULT  44
AC064835
ID   AC064835   standard; DNA; HTG; 202828 BP.
XX
AC   AC064835;
XX
SV   AC064835.3
XX
DT   26-APR-2000 (Rel. 63, Created)
DT   10-JUL-2000 (Rel. 64, Last updated, Version 6)
XX
DE   Homo sapiens chromosome 16 clone RP11-673P17, WORKING DRAFT SEQUENCE, 12
DE   unordered pieces.
XX
KW   HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-202828
```

```
RA   Waterston R.H.;
RT   "The sequence of Homo sapiens clone";
RL   Unpublished.
XX
RN   [2]
RP   1-202828
RA   Waterston R.H.;
RT   ;
RL   Submitted (22-APR-2000) to the EMBL/GenBank/DDBJ databases.
RL   Genome Sequencing Center, Washington University School of Medicine, 4444
RL   Forest Park Parkway, St. Louis, MO 63108, USA
XX
CC   On May 16, 2000 this sequence version replaced gi:7712280.
CC   ---------------- Genome Center ----------------
CC   Center: Washington University Genome Sequencing Center
CC   Center code: WUGSC
CC   Web site:http://genome.wustl.edu/gsc/index.shtml
CC   -------------- Project Information --------------
CC   Center project name: H_NH0673P17
CC   -------------- Summary Statistics --------------
CC   Sequencing vector: M13; 100%
CC   Sequencing vector: plasmid; 0%
CC   Chemistry: Dye-primer ET; 100% of reads
CC   Chemistry: Dye-terminator Big Dye; 0% of reads
CC   Assembly program: Phrap; version 0.990319
CC   Consensus quality: 195107 bases at least Q40
CC   Consensus quality: 197538 bases at least Q30
CC   Consensus quality: 198979 bases at least Q20
CC   Insert size: 199000; agarose-fp
CC   Insert size: 201728; sum-of-contigs
CC   Quality coverage: 5.33 in Q20 bases; agarose-fp
CC   Quality coverage: 5.29 in Q20 bases; sum-of-contigs
CC   ------------------------------------------------.
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 12 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1     1109: contig of 1109 bp in length
CC   *     1110     1209: gap of unknown length
CC   *     1210     5935: contig of 4726 bp in length
CC   *     5936     6035: gap of unknown length
CC   *     6036    13020: contig of 6985 bp in length
CC   *    13021    13120: gap of unknown length
CC   *    13121    21028: contig of 7908 bp in length
CC   *    21029    21128: gap of unknown length
CC   *    21129    27413: contig of 6285 bp in length
CC   *    27414    27513: gap of unknown length
CC   *    27514    34551: contig of 7038 bp in length
CC   *    34552    34651: gap of unknown length
CC   *    34652    53143: contig of 18492 bp in length
CC   *    53144    53243: gap of unknown length
CC   *    53244    77201: contig of 23958 bp in length
CC   *    77202    77301: gap of unknown length
CC   *    77302   106156: contig of 28855 bp in length
CC   *   106157   106256: gap of unknown length
CC   *   106257   131644: contig of 25388 bp in length
CC   *   131645   131744: gap of unknown length
CC   *   131745   158698: contig of 26954 bp in length
CC   *   158699   158798: gap of unknown length
CC   *   158799   202828: contig of 44030 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .202828
FT                   /chromosome="16"
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /clone="RP11-673P17"
FT   misc_feature    1. .1109
FT                   /note="assembly_name:Contig14"
FT   misc_feature    1210. .5935
FT                   /note="assembly_name:Contig18"
FT   misc_feature    6036. .13020
FT                   /note="assembly_name:Contig19"
FT   misc_feature    13121. .21028
FT                   /note="assembly_name:Contig20"
FT   misc_feature    21129. .27413
FT                   /note="assembly_name:Contig21 clone_end:T7
FT                   vector_side:right"
FT   misc_feature    27514. .34551
FT                   /note="assembly_name:Contig22"
FT   misc_feature    34652. .53143
FT                   /note="assembly_name:Contig23"
FT   misc_feature    53244. .77201
FT                   /note="assembly_name:Contig24"
FT   misc_feature    77302. .106156
FT                   /note="assembly_name:Contig25"
FT   misc_feature    106257. .131644
FT                   /note="assembly_name:Contig26 clone_end:SP6
FT                   vector_side:left"
FT   misc_feature    131745. .158698
FT                   /note="assembly_name:Contig27"
FT   misc_feature    158799. .202828
FT                   /note="assembly_name:Contig28"
XX
SQ   Sequence 202828 BP; 53011 A; 46987 C; 47342 G; 54379 T; 1109 other;
```

```
  Query Match             64.0%;  Score 16;  DB 33;  Length 202828;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db   46988 CAGGTGCAGGTCAGCA 47003


RESULT  45
AC016186/c
LOCUS       AC016186   245499 bp    DNA             HTG       12-MAR-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-8I15 map 18, *** SEQUENCING
            IN PROGRESS ***, 70 unordered pieces.
ACCESSION   AC016186
VERSION     AC016186.3  GI:7230177
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 245499)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-8I15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245499)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6649364.
```

```
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L4258
    Center clone name: 8_I_15
--------------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1174: contig of 1174 bp in length
*     1175 1274: gap of      100 bp
*     1275     2609: contig of 1335 bp in length
*     2610 2709: gap of      100 bp
*     2710     3931: contig of 1222 bp in length
*     3932 4031: gap of      100 bp
*     4032     5121: contig of 1090 bp in length
*     5122 5221: gap of      100 bp
*     5222     6287: contig of 1066 bp in length
*     6288 6387: gap of      100 bp
*     6388     7686: contig of 1299 bp in length
*     7687 7786: gap of      100 bp
*     7787     8797: contig of 1011 bp in length
*     8798 8897: gap of      100 bp
*     8898    10536: contig of 1639 bp in length
*    10537 10636: gap of      100 bp
*    10637    11858: contig of 1222 bp in length
*    11859 11958: gap of      100 bp
*    11959    13375: contig of 1417 bp in length
*    13376 13475: gap of      100 bp
*    13476    15102: contig of 1627 bp in length
*    15103 15202: gap of      100 bp
*    15203    16445: contig of 1243 bp in length
*    16446 16545: gap of      100 bp
*    16546    17637: contig of 1092 bp in length
*    17638 17737: gap of      100 bp
*    17738    19309: contig of 1572 bp in length
*    19310 19409: gap of      100 bp
*    19410    21532: contig of 2123 bp in length
*    21533 21632: gap of      100 bp
*    21633    23166: contig of 1534 bp in length
*    23167 23266: gap of      100 bp
*    23267    25262: contig of 1996 bp in length
*    25263 25362: gap of      100 bp
*    25363    27315: contig of 1953 bp in length
*    27316 27415: gap of      100 bp
*    27416    28722: contig of 1307 bp in length
*    28723 28822: gap of      100 bp
*    28823    30596: contig of 1774 bp in length
*    30597 30696: gap of      100 bp
*    30697    33203: contig of 2507 bp in length
*    33204 33303: gap of      100 bp
*    33304    35408: contig of 2105 bp in length
*    35409 35508: gap of      100 bp
*    35509    37441: contig of 1933 bp in length
*    37442 37541: gap of      100 bp
*    37542    40401: contig of 2860 bp in length
*    40402 40501: gap of      100 bp
*    40502    42052: contig of 1551 bp in length
*    42053 42152: gap of      100 bp
*    42153    44294: contig of 2142 bp in length
*    44295 44394: gap of      100 bp
*    44395    46328: contig of 1934 bp in length
*    46329 46428: gap of      100 bp
*    46429    48649: contig of 2221 bp in length
*    48650 48749: gap of      100 bp
*    48750    51037: contig of 2288 bp in length
*    51038 51137: gap of      100 bp
*    51138    53348: contig of 2211 bp in length
*    53349 53448: gap of      100 bp
*    53449    55337: contig of 1889 bp in length
*    55338 55437: gap of      100 bp
*    55438    57028: contig of 1591 bp in length
*    57029 57128: gap of      100 bp
*    57129    59830: contig of 2702 bp in length
*    59831 59930: gap of      100 bp
*    59931    62172: contig of 2242 bp in length
*    62173 62272: gap of      100 bp
*    62273    64940: contig of 2668 bp in length
*    64941 65040: gap of      100 bp
*    65041    67862: contig of 2822 bp in length
*    67863 67962: gap of      100 bp
*    67963    70236: contig of 2274 bp in length
*    70237 70336: gap of      100 bp
*    70337    72533: contig of 2197 bp in length
*    72534 72633: gap of      100 bp
*    72634    74796: contig of 2163 bp in length
*    74797 74896: gap of      100 bp
*    74897    77899: contig of 3003 bp in length
*    77900 77999: gap of      100 bp
*    78000    81075: contig of 3076 bp in length
*    81076 81175: gap of      100 bp
*    81176    84346: contig of 3171 bp in length
*    84347 84446: gap of      100 bp
*    84447    87026: contig of 2580 bp in length
*    87027 87126: gap of      100 bp
*    87127    90859: contig of 3733 bp in length
*    90860 90959: gap of      100 bp
*    90960    93599: contig of 2640 bp in length
*    93600 93699: gap of      100 bp
*    93700    96946: contig of 3247 bp in length
*    96947 97046: gap of      100 bp
*    97047   100076: contig of 3030 bp in length
*   100077 100176: gap of      100 bp
*   100177   105333: contig of 5157 bp in length
*   105334 105433: gap of      100 bp
*   105434   109169: contig of 3736 bp in length
*   109170 109269: gap of      100 bp
*   109270   113940: contig of 4671 bp in length
*   113941 114040: gap of      100 bp
*   114041   119295: contig of 5255 bp in length
*   119296 119395: gap of      100 bp
*   119396   120878: contig of 1483 bp in length
*   120879 120978: gap of      100 bp
*   120979   123579: contig of 2601 bp in length
*   123580 123679: gap of      100 bp
*   123680   128211: contig of 4532 bp in length
*   128212 128311: gap of      100 bp
*   128312   132536: contig of 4225 bp in length
*   132537 132636: gap of      100 bp
*   132637   137226: contig of 4590 bp in length
*   137227 137326: gap of      100 bp
*   137327   142521: contig of 5195 bp in length
*   142522 142621: gap of      100 bp
*   142622   147070: contig of 4449 bp in length
*   147071 147170: gap of      100 bp
*   147171   151866: contig of 4696 bp in length
*   151867 151966: gap of      100 bp
*   151967   156914: contig of 4948 bp in length
*   156915 157014: gap of      100 bp
*   157015   162849: contig of 5835 bp in length
*   162850 162949: gap of      100 bp
*   162950   169655: contig of 6706 bp in length
*   169656 169755: gap of      100 bp
*   169756   174040: contig of 4285 bp in length
*   174041 174140: gap of      100 bp
```

```
            *   174141   180979: contig of 6839 bp in length
            *   180980 181079: gap of      100 bp
            *   181080   190751: contig of 9672 bp in length
            *   190752 190851: gap of      100 bp
            *   190852   202044: contig of 11193 bp in length
            *   202045 202144: gap of      100 bp
            *   202145   211327: contig of 9183 bp in length
            *   211328 211427: gap of      100 bp
            *   211428   224152: contig of 12725 bp in length
            *   224153 224252: gap of      100 bp
            *   224253   234236: contig of 9984 bp in length
            *   234237 234336: gap of      100 bp
            *   234337   245499: contig of 11163 bp in length.
FEATURES             Location/Qualifiers
     source          1. .245499
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-8I15"
                     /clone_lib="RPCI-11 Human Male BAC"

  Query Match             64.0%;  Score 16;  DB 2;  Length 245499;
  Best Local Similarity   100.0%;  Pred. No. 58;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db 109551 CAGGTGCAGGTCAGCA 109536


Search completed: December 26, 2001, 12:54:24
Job time: 8986 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:17:08 ; Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 47864

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25	100.0	839	22	AAI62941	Human genomic DNA
2		25	100.0	1041	20	AAX01303	E. coli biotin syn
3		25	100.0	1084	10	AAN91329	E.coli Bio B gene.
4		25	100.0	1121	7	AAN60496	Sequence encoding
5		25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
6	c	16	64.0	450	22	AAI12043	Probe #1976 for ge
7	c	16	64.0	450	22	AAI33378	Probe #2064 used t
8	c	16	64.0	450	22	AAI01968	Probe #1959 used t
9	c	15	60.0	353	21	AAC56641	Eucalyptus grandis
10	c	15	60.0	630	16	AAQ78948	Human immunoglobul
11	c	15	60.0	1969	22	AAI14105	Probe #4038 for ge
12	c	15	60.0	1969	22	AAI35486	Probe #4172 used t
13	c	15	60.0	1969	22	AAI03958	Probe #3949 used t
14	c	15	60.0	4674	20	AAX90421	Human ataxin-2 lik
15		14	56.0	269	21	AAA31747	Plant microsatelli
16	c	14	56.0	327	21	AAC00048	Human secreted pro
17	c	14	56.0	363	22	AAH42390	Nucleotide sequenc
18	c	14	56.0	374	22	AAF64981	Novel human polynu
19	c	14	56.0	423	14	AAQ45597	Sequence encoding
20	c	14	56.0	423	20	AAZ39427	Antibody B17X2 VH
21	c	14	56.0	423	20	AAZ23965	Human B17X2 antibo
22	c	14	56.0	444	18	AAT80181	Monoclonal antibod
23	c	14	56.0	447	20	AAZ20406	IgG antibody 2.4.4
24	c	14	56.0	613	16	AAQ78943	Human immunoglobul
25	c	14	56.0	981	21	AAA15905	Human protein clon
26	c	14	56.0	1100	21	AAA15915	Human protein clon
27	c	14	56.0	1130	21	AAA39064	Human secreted pro
28	c	14	56.0	1135	21	AAA39087	Human secreted pro
29	c	14	56.0	1161	22	AAA91019	Human secreted pro
30	c	14	56.0	1173	21	AAZ98136	Human signal pepti
31	c	14	56.0	1186	21	AAC98064	Human colon cancer
32	c	14	56.0	1186	22	AAH35018	Human colon cancer
33		14	56.0	1191	19	AAT98726	DNA encoding a S.
34	c	14	56.0	1368	22	AAF44623	Novel protein kina
35		14	56.0	1634	19	AAZ96242	S. pneumoniae deri
36		14	56.0	1634	19	AAV37360	Streptococcus pneu
37	c	14	56.0	1814	22	AAI59259	Human polynucleoti
38		14	56.0	1816	22	AAI61045	Human polynucleoti
39		14	56.0	37856	21	AAA11992	S. cellulosum DNA
40		14	56.0	534720	19	AAV30458	Rhizobium species
41		14	56.0	536165	19	AAV30459	Rhizobium species
42		13	52.0	18	12	AAQ13647	Exon B/intron junc
43	c	13	52.0	123	20	AAX85925	Oligonucleotide us
44		13	52.0	153	22	AAI23772	Probe #13705 for g
45		13	52.0	153	22	AAI49080	Probe #17766 used
46		13	52.0	153	22	AAI09377	Probe #9368 used t
47	c	13	52.0	225	14	AAQ60177	Human brain Expres
48	c	13	52.0	269	22	AAI22207	Probe #12140 for g
49	c	13	52.0	276	21	AAC00049	Human secreted pro
50		13	52.0	279	22	AAI26888	Probe #16821 for g
51		13	52.0	279	22	AAI55684	Probe #24370 used
52	c	13	52.0	363	20	AAX85929	DNA encoding the h
53	c	13	52.0	363	20	AAX79527	Heavy chain coding
54	c	13	52.0	370	22	AAH34737	Human colon cancer
55		13	52.0	375	21	AAF22004	Human breast and o
56	c	13	52.0	378	18	AAT61237	Human anti-RSV mon
57	c	13	52.0	378	18	AAT61239	Human anti-RSV mon
58		13	52.0	386	22	AAI14576	Probe #4509 for ge
59		13	52.0	386	22	AAI35944	Probe #4630 used t
60		13	52.0	386	22	AAI04386	Probe #4377 used t
61	c	13	52.0	395	22	AAI12969	Probe #2902 for ge
62	c	13	52.0	400	21	AAC00033	Human secreted pro
63	c	13	52.0	423	20	AAX79207	Anti-HIV-1 gp120 a
64	c	13	52.0	437	21	AAZ42289	Human 5' EST isola
65		13	52.0	463	22	AAI16085	Probe #6018 for ge
66		13	52.0	463	22	AAI38303	Probe #6989 used t
67		13	52.0	472	21	AAC75419	Human ORFX ORF974
68	c	13	52.0	477	20	AAZ20408	IgG antibody 2.3.2
69		13	52.0	480	22	AAH65984	C glutamicum codin
70		13	52.0	527	20	AAV87894	EST clone FG380.
71	c	13	52.0	529	22	AAH13160	Human cDNA clone (
72	c	13	52.0	546	16	AAQ78964	Human immunoglobul
73		13	52.0	553	22	AAH97949	Murine 7-transmemb
74		13	52.0	588	22	AAI17685	Probe #7618 for ge
75		13	52.0	588	22	AAI42619	Probe #11305 used
76		13	52.0	593	22	AAI40100	Probe #8786 used t
77	c	13	52.0	600	22	AAH66012	C glutamicum codin
78	c	13	52.0	699	21	AAA02409	Human colon cancer
79		13	52.0	718	22	AAA89146	Polyketide synthas
80	c	13	52.0	723	22	AAF70988	C. glutamicum SRT
81	c	13	52.0	732	22	AAH70228	Human cervical can
82		13	52.0	750	20	AAX58277	Zea mays SCLBr pro
83	c	13	52.0	879	16	AAT01341	Human FK506 bindin
84	c	13	52.0	880	21	AAF20929	Human FK-506 bindi

c	85	13	52.0	880	21	AAA34807	Human adenosine re
	86	13	52.0	911	22	AAH29728	S cerevisiae apopt
c	87	13	52.0	933	22	AAH32118	Human olfactory re
	88	13	52.0	939	22	AAH67432	C glutamicum codin
c	89	13	52.0	952	22	AAH32030	Human olfactory re
c	90	13	52.0	963	22	AAH67430	C glutamicum codin
c	91	13	52.0	969	22	AAF28571	Fragment of pyciKC
	92	13	52.0	1040	20	AAX58275	Zea mays SCLBr pro
c	93	13	52.0	1101	19	AAV34298	Human secreted pro
	94	13	52.0	1113	22	AAF71611	Corynebacterium gl
	95	13	52.0	1121	22	AAD07933	Human TDC protein-
c	96	13	52.0	1122	21	AAZ98342	A. thaliana gene i
	97	13	52.0	1160	21	AAA29322	Glycine max isofla
c	98	13	52.0	1173	22	AAC68663	Human D1B2 cDNA.
c	99	13	52.0	1215	22	AAH45073	Murine FM-3 coding
c	100	13	52.0	1282	22	AAF28572	Fragment of pyciKC

ALIGNMENTS

RESULT 1
AAI62941
ID AAI62941 standard; DNA; 839 BP.
XX
AC AAI62941;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human genomic DNA SEQ ID NO 269.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200155449-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01346.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0218290.
PR 14-AUG-2000; 2000US-0225447.
PR 01-SEP-2000; 2000US-0229343.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231243.
PR 25-SEP-2000; 2000US-0234997.
PR 29-SEP-2000; 2000US-0236367.
PR 13-OCT-2000; 2000US-0239937.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246528.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249265.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476225/51.
XX
PT Novel plasma membrane associated proteins useful for diagnosing,
PT treating, preventing and/or prognosing disorders related to the
PT proteins, including cancer, immune response and neuronal disorders -
XX
PS Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    143 tcgtcaggtgcaggtcagcacgttg 167
```

RESULT 2
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX

```
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Patton DA;
XX
DR   WPI; 1999-152902/13.
DR   P-PSDB; AAW73906.
XX
PT   Transgenic plants with high biotin levels - transformed with DNA
PT   encoding di:amino-pelargonic acid amino-transferase or biotin
PT   synthase
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.00035;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     111 tcgtcaggtgcaggtcagcacgttg 135


RESULT   3
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA:
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.00035;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     134 tcgtcaggtgcaggtcagcacgttg 158


RESULT   4
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.00035;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     152 tcgtcaggtgcaggtcagcacgttg 176


RESULT   5
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
```

```
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.00032;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 tcgtcaggtgcaggtcagcacgttg 251


RESULT   6
AAI12043/c
ID   AAI12043 standard; DNA; 450 BP.
XX
AC   AAI12043;
```

```
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #1976 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 1976; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 450 BP; 121 A; 117 C; 124 G; 88 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 22;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db     423 GTCAGGTGCAGGTCAG 408


RESULT   7
AAI33378/c
ID   AAI33378 standard; DNA; 450 BP.
XX
AC   AAI33378;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #2064 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 2064; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 450 BP; 121 A; 117 C; 124 G; 88 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 22;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db     423 GTCAGGTGCAGGTCAG 408


RESULT   8
AA101968/c
ID   AA101968 standard; DNA; 450 BP.
XX
AC   AA101968;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Probe #1959 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
```

```
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 1959; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 450 BP; 121 A; 117 C; 124 G; 88 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 22;  Length 450;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db     423 GTCAGGTGCAGGTCAG 408


RESULT   9
AAC56641/c
ID   AAC56641 standard; DNA; 353 BP.
XX
AC   AAC56641;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #512.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Page 482; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 353 BP; 117 A; 59 C; 98 G; 78 T; 1 other;


  Query Match             60.0%;  Score 15;  DB 21;  Length 353;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggt 15
           |||||||||||||||
Db     326 TCGTCAGGTGCAGGT 312


RESULT  10
AAQ78948/c
ID   AAQ78948 standard; DNA; 630 BP.
XX
AC   AAQ78948;
XX
DT   01-AUG-1995  (first entry)
XX
DE   Human immunoglobulin Vh gene #10.
XX
KW   Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;
KW   cosmid; placenta; vector; pJB81; E.coli; mammalian; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             71..495
FT                   /*tag=  a
FT                   /product= human immunoglobulin variable heavy chain
FT   intron          114..199
FT                   /*tag=  b
FT   misc_signal     322..324
FT                   /*tag=  c
FT                   /transl_except= unused termination codon
FT   misc_signal     373..375
FT                   /*tag=  d
FT                   /transl_except= unused termination codon
FT   misc_signal     496..498
FT                   /*tag=  e
FT                   /note= "miscellaneous signal, does not conform to
FT                         termination or splice site sequence"
XX
PN   WO9426895-A.
XX
PD   24-NOV-1994.
XX
PF   10-MAY-1993;   93WO-JP00603.
XX
PR   10-MAY-1993;   93WO-JP00603.
XX
PA   (NISB ) JAPAN TOBACCO INC.
XX
PI   Honjo T,  Matsuda F;
XX
```

DR WPI; 1995-006791/01.
DR P-PSDB; AAR66304.
XX
PT DNA fragment comprising human immunoglobulin Vh genes - for the
PT production of human immunoglobulin in mammalian hosts
XX
PS Claim 19; Page 43-44; 130pp; Japanese.
XX
CC A series of genes (AAQ78939-79002) encoding human immunoglobulin
CC variable heavy chains. The genes were isolated and cloned from a series
CC of cosmid constructs: Y202; Y103; Y21; Y6;Y24; 3-31; M84; M118 and M131,
CC by PCR amplification using primers AAQ78917-38. The genes are subdivided
CC into 5 families of Vh genes. The fragments cover a region of 800 kb.
CC The DNA fragments were isolated from high molecular weight DNA from
CC human placenta. The DNA was partially digested with TaqI restriction
CC enzyme. The fragments were separated by gel elecrophoresis and 35-45 kb
CC fractions were collected. The fragments were ligated with ClaI-digested
CC cosmid vector pJB81. The ligation products were in vitro packed and
CC infected into E.coli 490A. The fragments were then subcloned by colony
CC hybridisation. The Vh genes and the DNA fragments encoding them are
CC useful in producing human immunoglobulin in mammalian hosts.
XX
SQ Sequence 630 BP; 141 A; 179 C; 148 G; 162 T; 0 other;

Query Match 60.0%; Score 15; DB 16; Length 630;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagca 20
||||||||||||||||
Db 280 AGGTGCAGGTCAGCA 266

RESULT 11
AAI14105/c
ID AAI14105 standard; DNA; 1969 BP.
XX
AC AAI14105;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #4038 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 4038; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1969 BP; 551 A; 432 C; 640 G; 346 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 1969;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 tgcaggtcagcacgt 23
|||||||||||||||
Db 1418 TGCAGGTCAGCACGT 1404

RESULT 12
AAI35486/c
ID AAI35486 standard; DNA; 1969 BP.
XX
AC AAI35486;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #4172 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 4172; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 1969 BP; 551 A; 432 C; 640 G; 346 T; 0 other;

```
  Query Match            60.0%;  Score 15;  DB 22;  Length 1969;
  Best Local Similarity  100.0%;  Pred. No. 37;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 tgcaggtcagcacgt 23
          |||||||||||||||
Db   1418 TGCAGGTCAGCACGT 1404


RESULT  13
AAI03958/c
ID   AAI03958 standard; DNA; 1969 BP.
XX
AC   AAI03958;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Probe #3949 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 3949; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1969 BP; 551 A; 432 C; 640 G; 346 T; 0 other;


  Query Match            60.0%;  Score 15;  DB 22;  Length 1969;
  Best Local Similarity  100.0%;  Pred. No. 37;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 tgcaggtcagcacgt 23
          |||||||||||||||
Db   1418 TGCAGGTCAGCACGT 1404


RESULT  14
AAX90421/c
ID   AAX90421 standard; cDNA; 4674 BP.
XX
AC   AAX90421;
XX
DT   29-SEP-1999  (first entry)
XX
DE   Human ataxin-2 like gene.
XX
KW   Human; ataxin-2 like gene; ataxin-2 like protein; ataxia; deafness;
KW   cardiomyopathy; neurological disease; cancer; AIDS; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             63..3218
FT                   /*tag=  a
FT                   /product= "ataxin-2 like protein"
XX
PN   WO9936527-A1.
XX
PD   22-JUL-1999.
XX
PF   19-JAN-1998;   98WO-CN00009.
XX
PR   19-JAN-1998;   98WO-CN00009.
XX
PA   (UYHU-) UNIV HUNAN MEDICAL.
XX
PI   Deng H,  Liu C,  Wang D,  Xia J;
XX
DR   WPI; 1999-458463/38.
DR   P-PSDB; AAY29321.
XX
PT   Ataxin-2 like protein, and related polynucleotides, useful in
PT   treatment and diagnosis of ataxia,
XX
PS   Claim 4; Page 21-23; 34pp; English.
XX
CC   The present sequence represents a human ataxin-2 like gene. Ataxin-2
CC   like polynucleotides (I), vectors containing (I) and recombinant host
CC   cells are useful for recombinant production of Ataxin-2 like protein.
CC   Both (I), ataxin-2 like protein and antibodies against ataxin-2 like
CC   protein are useful as research reagents, for screening assays and in
CC   diagnostic assays. Antagonists and agonists of ataxin-2 like protein
CC   can be used to inhibit or enhance, respectively, the activity of
CC   ataxin-2 like protein or expression of (I). Anti-ataxin-2 like protein
CC   antibodies and ataxin-2 like protein or its fragments can be used
CC   in vaccines. In particular, the proteins, antibodies, agonists and
CC   antagonists can be used for treating, e.g. ataxia, cardiomyopathy,
CC   deafness, neurological disease, cancer and AIDS, related to both an
CC   excess and insufficient amounts of ataxin-2 like.
XX
SQ   Sequence 4674 BP; 971 A; 1566 C; 1244 G; 893 T; 0 other;


  Query Match            60.0%;  Score 15;  DB 20;  Length 4674;
  Best Local Similarity  100.0%;  Pred. No. 36;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagc 19
          |||||||||||||||
Db   3741 CAGGTGCAGGTCAGC 3727


RESULT  15
AAA31747
ID   AAA31747 standard; DNA; 269 BP.
XX
AC   AAA31747;
```

XX
DT 05-JUL-2000 (first entry)
XX
DE Plant microsatellite marker #708.
XX
KW Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.
XX
OS Eucalyptus grandis.
XX
PN WO9967421-A1.
XX
PD 29-DEC-1999.
XX
PF 25-JUN-1999; 99WO-NZ00092.
XX
PR 25-JUN-1998; 98US-0105307.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Havukkala IJ, Bloksberg LN, Glenn M;
XX
DR WPI; 2000-116958/10.
XX
PT New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -
XX
PS Claim 1; Page 285; 392pp; English.
XX
CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences AAA32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
SQ Sequence 269 BP; 63 A; 62 C; 61 G; 83 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 caggtcagcacgtt 24
||||||||||||||
Db 76 caggtcagcacgtt 89

RESULT 16
AAC00048/c
ID AAC00048 standard; cDNA; 327 BP.
XX
AC AAC00048;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 46.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG00042.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 46; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 327 BP; 70 A; 102 C; 75 G; 69 T; 11 other;

Query Match 56.0%; Score 14; DB 21; Length 327;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 144 AGGTGCAGGTCAGC 131

RESULT 17
AAH42390/c
ID AAH42390 standard; DNA; 363 BP.
XX
AC AAH42390;
XX
DT 01-OCT-2001 (first entry)
XX
DE Nucleotide sequence of variable heavy chain fragment of clone G81.
XX
KW Antibody; heavy chain; VH; amyloid protein; blood brain barrier;
KW endothelial cell; brain cell antigen; inflammation; adhesion molecule;
KW transferrin receptor; neurological disease; Alzheimer's disease;
KW prion disease; AIDS-related dementia; epilepsy; brain injury; ss.
XX
OS Homo sapiens.
XX
PN WO200144300-A2.
XX
PD 21-JUN-2001.
XX
PF 27-NOV-2000; 2000WO-GB04501.
XX
PR 13-DEC-1999; 99US-0170599.
XX
PA (CAMB-) CAMBRIDGE ANTIBODY TECHNOLOGY.
XX

PI Webster C, Osbourn J, Ward G, Miller K;
XX
DR WPI; 2001-398131/42.
DR P-PSDB; AAG62946.
XX
PT Mixture or panel of antibodies for selecting specific binding members
PT that cross the blood brain barrier, for use in delivering different
PT molecules and treating neurological diseases -
XX
PS Disclosure; Page 98; 109pp; English.
XX
CC The present sequence encodes an antibody variable heavy chain (VH)
CC fragment. The fragment is used to produce a mixture or panel of 5
CC different specific binding members, each comprising an antibody VH
CC and/or VL variable domain and capable, when displayed on the surface
CC of filamentous bacteriophage particles or in the case of a specific
CC binding member comprising the D5 VH and/or VL variable domain when
CC bound to human serum amyloid protein, to pass through a mammalian
CC blood brain barrier (BBB). The panel is useful for the selection of
CC specific binding members with a desired property such as ability to
CC cross BBB, ability to bind endothelial cells or other brain cell antigen,
CC ability to bind areas of inflammation in the brain or BBB breakdown or
CC ability to bind intracellular adhesion molecules and to bind transferrin
CC receptor. The antibodies are useful in diagnosis, prophylaxis and
CC treatment of human or animal body, including neurological diseases, such
CC as Alzheimer's disease, prion disease, AIDS-related dementia, epilepsy
CC and traumatic brain injury and any diseases involving inflammation
CC occurring within the brain or central nervous system.
XX
SQ Sequence 363 BP; 84 A; 108 C; 96 G; 75 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 363;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 70 AGGTGCAGGTCAGC 57

RESULT 18
AAF64981/c
ID AAF64981 standard; cDNA; 374 BP.
XX
AC AAF64981;
XX
DT 09-APR-2001 (first entry)
XX
DE Novel human polynucleotide, SEQ ID NO: 737.
XX
KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.
XX
OS Homo sapiens.
XX
PN WO200102568-A2.
XX
PD 11-JAN-2001.
XX
PF 30-JUN-2000; 2000WO-US18374.
XX
PR 02-JUL-1999; 99US-0142310.
PR 02-JUL-1999; 99US-0142311.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitiz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 650; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 374 BP; 71 A; 116 C; 118 G; 68 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 180 AGGTGCAGGTCAGC 167

RESULT 19
AAQ45597/c
ID AAQ45597 standard; DNA; 423 BP.
XX
AC AAQ45597;
XX
DT 04-DEC-1993 (first entry)
XX
DE Sequence encoding the VH of antibody B17X2.
XX
KW Variable heavy antibody chain; human subgroup 4 germline; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 7..423
FT /*tag= a
XX
PN WO9312231-A.
XX
PD 24-JUN-1993.
XX
PF 13-DEC-1991; 91WO-AU00583.
XX
PR 13-DEC-1991; 91WO-AU00583.
XX
PA (DOWC) DOW CHEM AUSTRALIA LTD.
XX
PI Johnson KS, Mezes PS, Richard RA;
XX
DR WPI; 1993-214173/26.
DR P-PSDB; AAR38315.
XX
PT New composite antibody binding to tumour associated TAG-72
PT antigen - includes light chain variable region from human
PT subgroup 4 germline gene, useful, opt. as conjugate, for
PT diagnosis or treatment of cancer

```
XX
PS   Disclosure; Figure 4; 150pp; English.
XX
CC   Cell line B17X2 expresses an antibody utilising a variable light
CC   chain encoded by a gene derived from Hum4 VL and a variable heavy
CC   chain which makes a stable VL and VH combination.
XX
SQ   Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;


  Query Match             56.0%;  Score 14;  DB 14;  Length 423;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     133 AGGTGCAGGTCAGC 120


RESULT  20
AAZ39427/c
ID   AAZ39427 standard; DNA; 423 BP.
XX
AC   AAZ39427;
XX
DT   29-FEB-2000  (first entry)
XX
DE   Antibody B17X2 VH nucleotide sequence.
XX
KW   Tumor associated sialylated glycoprotein; TAG-72; cancer antigen;
KW   carcinoma lesion; diagnostic; cancer; antibody; human; B17X2;
KW   anti-mouse antibody hypersensitivity reaction; ss.
XX
OS   Homo sapiens.
XX
PN   US5976845-A.
XX
PD   02-NOV-1999.
XX
PF   07-JUN-1995;   95US-0487743.
XX
PR   16-JUN-1994;   94US-0261354.
PR   19-APR-1990;   90US-0510697.
PR   20-OCT-1992;   92US-0964536.
XX
PA   (DOWC ) DOW CHEM CO.
XX
PI   Johnson KS,  Richard RA,  Mezes PS;
XX
DR   WPI; 1999-619651/53.
DR   P-PSDB; AAY57179.
XX
PT   Production of humanized anti-TAG-72 antibodies, used for the detection,
PT   in vivo imaging and treatment of cancers  -
XX
PS   Disclosure; Fig 4A-B; 85pp; English.
XX
CC   The invention relates to producing humanized anti-tumor associated
CC   sialylated glycoprotein (TAG-72) antibodies (anti-TAG). The antibodies
CC   have binding specificity for the cancer antigen TAG-72. These antibodies
CC   have variable regions with VL segments derived from human subgroup IV
CC   germline gene and a VH segment (encoded by the VHalphaTAG germline gene)
CC   which is capapble of combining with the VL to form a threee dimensional
CC   structure having the ability to bind TAG-72. They can be used for the in
CC   vivo detection of carcinoma lesions. They can also be used for in vitro
CC   diagnostics. They can also be modified with therapeutic agents e.g. a
CC   radionuclide, drug, biological response modifier, toxin or another
CC   antibody for the treatment of cancers. The humanized anti-TAG-72
CC   antibodies can reduce harmful anti-mouse antibody hypersensitivity
CC   reactions.
XX
SQ   Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 423;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     133 AGGTGCAGGTCAGC 120


RESULT  21
AAZ23965/c
ID   AAZ23965 standard; DNA; 423 BP.
XX
AC   AAZ23965;
XX
DT   09-FEB-2000  (first entry)
XX
DE   Human B17X2 antibody VH segment DNA.
XX
KW   Human; antibody; humanized; anti-tumor; sialylated glycoprotein antibody;
KW   TAG-72; cytostatic; cancer antigen; detection; carcinoma lesion;
KW   diagnostic; treatment; ss.
XX
OS   Homo sapiens.
XX
PN   US5976531-A.
XX
PD   02-NOV-1999.
XX
PF   16-JUN-1994;   94US-0261354.
XX
PR   19-APR-1990;   90US-0510697.
PR   20-OCT-1992;   92US-0964536.
XX
PA   (DOWC ) DOW CHEM CO.
XX
PI   Johnson KS,  Mezes PS,  Richard RA;
XX
DR   WPI; 1999-632731/54.
DR   P-PSDB; AAY50688.
XX
PT   New humanized anti-TAG-72 antibodies, used for the detection, in vivo
PT   imaging and treatment of cancers  -
XX
PS   Disclosure; Figure 4A-B; 83pp; English.
XX
CC   This invention describes novel humanized anti-tumor associated
CC   sialylated glycoprotein antibodies (TAG-72) which have cytostatic
CC   activity. The antibodies have binding specificity for the cancer antigen
CC   TAG-72.  They can be used for the in vivo detection of carcinoma lesions.
CC   They can also be used for in vitro diagnostics. They can also be modified
CC   with therapeutic agents e.g. a radionuclide, drug, biological response
CC   modifier, toxin or another antibody for the treatment of cancers. The
CC   humanized anti-TAG-72 antibodies can reduce harmful anti-mouse antibody
CC   hypersensitivity reactions.
XX
SQ   Sequence 423 BP; 98 A; 120 C; 102 G; 97 T; 6 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 423;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     133 AGGTGCAGGTCAGC 120


RESULT  22
AAT80181/c
ID   AAT80181 standard; DNA; 444 BP.
XX
```

```
AC   AAT80181;
XX
DT   02-APR-1998  (first entry)
XX
DE   Monoclonal antibody 105AD7 heavy chain variable region DNA sequence.
XX
KW   Monoclonal antibody 105AD7; complementarity determining region;
KW   CDR; human; immune response; treatment; cancer; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..444
FT                   /*tag=  a
FT                   /transl_except= (pos: 157..159, aa: Gly)
FT                   /transl_except= (pos: 160..162, aa: Val)
FT                   /product= "105AD7 heavy chain variable region"
XX
PN   WO9732021-A1.
XX
PD   04-SEP-1997.
XX
PF   28-FEB-1997;   97WO-GB00591.
XX
PR   29-FEB-1996;   96GB-0004321.
PR   28-FEB-1996;   96GB-0004177.
XX
PA   (CANC-) CANCER RES CAMPAIGN TECHNOLOGY LTD.
XX
PI   Durrant LG,  Robins RA,  Spendlove I;
XX
DR   WPI; 1997-448689/41.
DR   P-PSDB; AAW26239.
XX
PT   Nucleic acid encoding peptide derived from antibody 105AD7 - used to
PT   induce an immune response for treatment and prevention of cancer
XX
PS   Claim 1; Fig 1A; 32pp; English.
XX
CC   This is the DNA sequence of the monoclonal antibody 105AD7 heavy chain
CC   variable region. 105AD7 is a human monoclonal anti-idiotypic antibody
CC   that mimics T cell epitopes on the tumour associated antigen gp72 and
CC   has been used for therapeutic vaccination. A peptide derived from this
CC   antibody contains the 105AD7 heavy chain variable region and a 105AD7
CC   kappa chain. The complementarity determining regions (CDR) of this
CC   peptide and its fragments can stimulate immune responses. The peptide
CC   can be used for screening for its functional equivalents and mimetics.
CC   The peptides, and the fragments, equivalents and mimetics are used to
CC   stimulate an immune response to a tumour antigen, for the treatment or
CC   prevention of tumours. Transfected host cells are used to produce this
CC   peptide and the vectors can be used to generate the peptide in vivo for
CC   stimulation of an immune response. The peptide contains promiscuous
CC   helper epitopes which stimulate a response against tumour antigens other
CC   than gp72. These epitopes may help a cytotoxic T lymphocyte response to
CC   any co-injected antigen.
XX
SQ   Sequence 444 BP; 100 A; 134 C; 107 G; 103 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 18;  Length 444;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     121 AGGTGCAGGTCAGC 108


RESULT  23
AAZ20406/c
ID   AAZ20406 standard; cDNA; 447 BP.
XX
AC   AAZ20406;
XX
DT   19-NOV-1999  (first entry)
XX
DE   IgG antibody 2.4.4 heavy chain coding sequence.
XX
KW   Antibody; CD147; IgM; ABX-CBL; activated T-cell killing; leukaemia;
KW   activated B-cell; monocyte; graft versus host disease; therapy; cancer;
KW   organ transplant rejection disease; lymphoma; pancreatic disease;
KW   autoimmune disease; inflammatory disease; arthritis; binding site; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..447
FT                   /*tag=  a
FT                   /note= "no stop codon given"
XX
PN   WO9945031-A2.
XX
PD   10-SEP-1999.
XX
PF   03-MAR-1999;   99WO-US04583.
XX
PR   03-MAR-1998;   98US-0034607.
PR   03-FEB-1999;   99US-0244253.
XX
PA   (ABGE-) ABGENIX INC.
XX
PI   Davis CG,  Blacher RW,  Corvalan JR,  Culwell AR,  Green LL,  Hales J;
PI   Havrilla N,  Ivanov VE,  Lipani JA,  Liu Q,  Weber RF,  Yang X;
XX
DR   WPI; 1999-540816/45.
DR   P-PSDB; AAY34305.
XX
PT   New monoclonal antibody, used for treating e.g. graft versus host
PT   disease, cancers, autoimmune diseases and inflammatory diseases  -
XX
PS   Disclosure; Fig 30; 245pp; English.
XX
CC   This sequence encodes the heavy chain of an antibody of the
CC   invention. The antibody is a monoclonal antibody (MAb) with an isotype
CC   that fixes complement and a variable region that binds to the epitope on
CC   CD147 bound by the IgM MAb ABX-CBL, providing that the antibody is not
CC   CBL1. The MAb can selectively kill activated T-cells, activated B-cells
CC   or resting or activated monocytes. The products and methods can be used
CC   for treating diseases involving activated T-cells or B-cells or
CC   monocytes, e.g. graft versus host disease (GVHD), organ transplant
CC   rejection diseases (e.g. renal transplant, ocular transplant), cancers
CC   (e.g. cancers of the blood (e.g. leukaemia's and lymphomas) and
CC   pancreatic), autoimmune diseases (e.g. lupus), and inflammatory diseases
CC   (e.g. arthritis).
XX
SQ   Sequence 447 BP; 96 A; 150 C; 106 G; 95 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 447;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db      34 AGGTGCAGGTCAGC 21


RESULT  24
AAQ78943/c
ID   AAQ78943 standard; DNA; 613 BP.
XX
AC   AAQ78943;
XX
DT   07-AUG-1995  (first entry)
XX
DE   Human immunoglobulin variable heavy chain #5.
```

```
XX
KW   Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;
KW   cosmid; placenta; vector; pJB81; E.coli; mammalian; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             71..513
FT                   /*tag=  a
FT                   /product= human immunoglobulin variable heavy chain
FT   intron          117..202
FT                   /*tag=  b
FT   misc_signal     514..516
FT                   /*tag=  c
FT                   /note= "miscellaneous signal, does not conform to
FT                          terminator or splice site sequence"
XX
PN   WO9426895-A.
XX
PD   24-NOV-1994.
XX
PF   10-MAY-1993;   93WO-JP00603.
XX
PR   10-MAY-1993;   93WO-JP00603.
XX
PA   (NISB ) JAPAN TOBACCO INC.
XX
PI   Honjo T,  Matsuda F;
XX
DR   WPI; 1995-006791/01.
DR   P-PSDB; AAR66299.
XX
PT   DNA fragment comprising human immunoglobulin Vh genes - for the
PT   production of human immunoglobulin in mammalian hosts
XX
PS   Disclosure; Page 36-37; 130pp; Japanese.
XX
CC   A series of genes (AAQ78939-79002) encoding human immunoglobulin
CC   variable heavy chains. The genes were isolated and cloned from a series
CC   of cosmid constructs: Y202; Y103; Y21; Y6;Y24; 3-31; M84; M118 and M131,
CC   by PCR amplification using primers AAQ78917-38.  The genes are subdivided
CC   into 5 families of Vh genes.  The fragments cover a region of 800 kb.
CC   The DNA fragments were isolated from high molecular weight DNA from
CC   human placenta.  The DNA was partially digested with TaqI restriction
CC   enzyme. The fragments were separated by gel elecrophoresis and 35-45 kb
CC   fractions were collected.  The fragments were ligated with ClaI-digested
CC   cosmid vector pJB81.  The ligation products were in vitro packed and
CC   infected into E.coli 490A.  The fragments were then subcloned by colony
CC   hybridisation.  The Vh genes and the DNA fragments encoding them are
CC   useful in producing human immunoglobulin in mammalian hosts.
XX
SQ   Sequence 613 BP; 139 A; 194 C; 139 G; 141 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 16;  Length 613;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     283 AGGTGCAGGTCAGC 270


RESULT  25
AAA15905/c
ID   AAA15905 standard; cDNA; 981 BP.
XX
AC   AAA15905;
XX
DT   12-JUN-2000  (first entry)
XX
DE   Human protein clone HP10568 coding sequence.
XX
KW   Human protein; hydrophobic domain; nutritional source; haematopoiesis;
KW   cytokine production; cell proliferation; cell differentiation;
KW   immune deficiency; infectious disease; autoimmune disorder; asthma;
KW   multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW   allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW   nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW   Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW   systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW   coagulation disorder; myocardial infarction; inflammatory condition;
KW   septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW   nephritis; therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200005367-A2.
XX
PD   03-FEB-2000.
XX
PF   22-JUL-1999;   99WO-JP03929.
XX
PR   24-JUL-1998;   98JP-0208820.
PR   07-AUG-1998;   98JP-0224105.
PR   25-AUG-1998;   98JP-0238116.
PR   09-SEP-1998;   98JP-0254736.
PR   29-SEP-1998;   98JP-0275505.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2000-182694/16.
DR   P-PSDB; AAY94857.
XX
PT   Novel human proteins having hydrophobic domains useful for treating
PT   osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
PT   multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
XX
PS   Claim 3; Page 189; 351pp; English.
XX
CC   This sequence encodes a human protein of the invention, which has
CC   hydrophobic domains. The DNA sequences can be used as a probe or as a
CC   genetic marker. The protein can also be used as a marker, and to identify
CC   potential genetic disorders. The DNA and protein can also be used as
CC   nutritional sources or supplements. The protein exhibits cytokine, cell
CC   proliferation, cell differentiation activities and induces production of
CC   other cytokines in certain cell populations. The protein also exhibits
CC   immune stimulating or immune suppressing activity. It can be used in the
CC   treatment of various immune deficiencies and disorders, and to treat
CC   infectious diseases caused by viral, bacterial, fungal or other
CC   infections. The protein is also used for treating autoimmune disorders
CC   such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
CC   arthritis. It is also useful in the treatment of allergic reactions and
CC   conditions such as asthma, and in immune suppression after organ
CC   transplantation. The protein is useful in regulation of haematopoiesis
CC   and consequently in the treatment of myeloid or lymphoid cell
CC   deficiencies. It is also used in compositions for tissue growth or
CC   regeneration. The protein is also used in the treatment of osteoporosis
CC   or osteoarthritis and in the treatment of periodontal disease and other
CC   tooth repair processes. The protein is used in the treatment of nervous
CC   system disorders such as Alzheimer's disease, Parkinson's disease, and
CC   Huntington's disease. They are useful for protection or regeneration and
CC   treatment of lung or liver fibrosis, reperfusion injury in various
CC   tissues, and conditions resulting from systemic cytokine damage. They are
CC   also used for promoting or inhibiting tissue differentiation. They are
CC   also used as contraceptives since they exhibit activin or inhibin related
CC   activities and as a fertility inducing therapeutic. They are used for
CC   treating various coagulation disorders and in treatment and prevention of
CC   conditions resulting from coagulation activities e.g. myocardial
CC   infarction or stroke. They also acts as receptors, receptor ligands or
CC   inhibitors or agonists of receptor/ligand interactions. They are used to
CC   treat inflammatory conditions such as septic shock, sepsis, ischaemia
CC   reperfusion injury, arthritis, and nephritis. They can be used to
```

```
CC   prevent tumours.
XX
SQ   Sequence 981 BP; 194 A; 299 C; 275 G; 213 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 981;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     139 AGGTGCAGGTCAGC 126


RESULT  26
AAA15915/c
ID   AAA15915 standard; cDNA; 1100 BP.
XX
AC   AAA15915;
XX
DT   12-JUN-2000  (first entry)
XX
DE   Human protein clone HP10568 full length coding sequence.
XX
KW   Human protein; hydrophobic domain; nutritional source; haematopoiesis;
KW   cytokine production; cell proliferation; cell differentiation;
KW   immune deficiency; infectious disease; autoimmune disorder; asthma;
KW   multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW   allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW   nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW   Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW   systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW   coagulation disorder; myocardial infarction; inflammatory condition;
KW   septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW   nephritis; therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200005367-A2.
XX
PD   03-FEB-2000.
XX
PF   22-JUL-1999;   99WO-JP03929.
XX
PR   24-JUL-1998;   98JP-0208820.
PR   07-AUG-1998;   98JP-0224105.
PR   25-AUG-1998;   98JP-0238116.
PR   09-SEP-1998;   98JP-0254736.
PR   29-SEP-1998;   98JP-0275505.
XX
PA   (SAGA ) SAGAMI CHEM RES CENT.
PA   (PROT-) PROTEGENE INC.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2000-182694/16.
DR   P-PSDB; AAY94857.
XX
PT   Novel human proteins having hydrophobic domains useful for treating
PT   osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
PT   multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
XX
PS   Claim 4; Page 203-205; 351pp; English.
XX
CC   This sequence encodes a human protein of the invention, which has
CC   hydrophobic domains. The DNA sequences can be used as a probe or as a
CC   genetic marker. The protein can also be used as a marker, and to identify
CC   potential genetic disorders. The DNA and protein can also be used as
CC   nutritional sources or supplements. The protein exhibits cytokine, cell
CC   proliferation, cell differentiation activities and induces production of
CC   other cytokines in certain cell populations. The protein also exhibits
CC   immune stimulating or immune suppressing activity. It can be used in the
CC   treatment of various immune deficiencies and disorders, and to treat
CC   infectious diseases caused by viral, bacterial, fungal or other
CC   infections. The protein is also used for treating autoimmune disorders
CC   such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
CC   arthritis. It is also useful in the treatment of allergic reactions and
CC   conditions such as asthma, and in immune suppression after organ
CC   transplantation. The protein is useful in regulation of haematopoiesis
CC   and consequently in the treatment of myeloid or lymphoid cell
CC   deficiencies. It is also used in compositions for tissue growth or
CC   regeneration. The protein is also used in the treatment of osteoporosis
CC   or osteoarthritis and in the treatment of periodontal disease and other
CC   tooth repair processes. The protein is used in the treatment of nervous
CC   system disorders such as Alzheimer's disease, Parkinson's disease, and
CC   Huntington's disease. They are useful for protection or regeneration and
CC   treatment of lung or liver fibrosis, reperfusion injury in various
CC   tissues, and conditions resulting from systemic cytokine damage. They are
CC   also used for promoting or inhibiting tissue differentiation. They are
CC   also used as contraceptives since they exhibit activin or inhibin related
CC   activities and as a fertility inducing therapeutic. They are used for
CC   treating various coagulation disorders and in treatment and prevention of
CC   conditions resulting from coagulation activities e.g. myocardial
CC   infarction or stroke. They also acts as receptors, receptor ligands or
CC   inhibitors or agonists of receptor/ligand interactions. They are used to
CC   treat inflammatory conditions such as septic shock, sepsis, ischaemia
CC   reperfusion injury, arthritis, and nephritis. They can be used to
CC   prevent tumours.
XX
SQ   Sequence 1100 BP; 221 A; 333 C; 314 G; 232 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 1100;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     195 AGGTGCAGGTCAGC 182


RESULT  27
AAA39064/c
ID   AAA39064 standard; cDNA; 1130 BP.
XX
AC   AAA39064;
XX
DT   30-AUG-2000  (first entry)
XX
DE   Human secreted protein gene 13 SEQ ID NO:23.
XX
KW   Human; secreted protein; cytostatic; anti-proliferative; vulnerary;
KW   immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
KW   hyperproliferative disorder; infectious disease; tissue regeneration;
KW   screening; food additive; preservative; wound healing;
KW   hyper-vascular disease; chromosome 11; ss.
XX
OS   Homo sapiens.
XX
PN   WO200017222-A1.
XX
PD   30-MAR-2000.
XX
PF   22-SEP-1999;   99WO-US22012.
XX
PR   23-SEP-1998;   98US-0101546.
PR   02-OCT-1998;   98US-0102895.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Rosen CA,  Duan RD,  Shi Y,  Lafleur DW,  Young PE,  Ni J;
PI   Komatsoulis G,  Endress GA,  Soppet DR;
XX
DR   WPI; 2000-283538/24.
DR   P-PSDB; AAB08903.
XX
```

```
PT   Human secreted proteins and coding sequences useful in diagnostic and
PT   therapeutic methods for disorders such as immune system or
PT   proliferative disorders, related to the proteins -
XX
PS   Claim 1; Page 329-330; 416pp; English.
XX
CC   The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC   human secreted proteins given in AAB08891 to AAB08984. The human secreted
CC   proteins can have activities based on the tissues and cells they are
CC   expressed in. Examples of the activities are: cytostatic;
CC   anti-proliferative; immunosuppressive; antibacterial; and vulnerary. The
CC   secreted proteins and their related polynucleotide sequences are useful
CC   for diagnostic and therapeutic methods useful for diagnosing and treating
CC   disorders related to the secreted proteins. The proteins, and
CC   polynucleotide sequences may be useful for treating disorders of the
CC   immune system, hyperproliferative disorders, infectious disease,
CC   regeneration of tissues, for chemotaxis and for screening molecules that
CC   bind to the proteins. The proteins or polynucleotide sequences may be
CC   used as food additives or preservatives, to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, co-factors or other nutritional components. Agonists or
CC   antagonists of the proteins may be used to prevent scar tissue growth
CC   during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
CC   and AAB08890 are sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1130 BP; 246 A; 334 C; 317 G; 233 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 1130;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    201 AGGTGCAGGTCAGC 188


RESULT  28
AAA39087/c
ID   AAA39087 standard; cDNA; 1135 BP.
XX
AC   AAA39087;
XX
DT   30-AUG-2000  (first entry)
XX
DE   Human secreted protein gene 13 SEQ ID NO:46.
XX
KW   Human; secreted protein; cytostatic; anti-proliferative; vulnerary;
KW   immunosuppressive; antibacterial; diagnosis; immune system; chemotaxis;
KW   hyperproliferative disorder; infectious disease; tissue regeneration;
KW   screening; food additive; preservative; wound healing;
KW   hyper-vascular disease; chromosome 11; ss.
XX
OS   Homo sapiens.
XX
PN   WO200017222-A1.
XX
PD   30-MAR-2000.
XX
PF   22-SEP-1999;   99WO-US22012.
XX
PR   23-SEP-1998;   98US-0101546.
PR   02-OCT-1998;   98US-0102895.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Rosen CA,  Duan RD,  Shi Y,  Lafleur DW,  Young PE,  Ni J;
PI   Komatsoulis G,  Endress GA,  Soppet DR;
XX
DR   WPI; 2000-283538/24.
DR   P-PSDB; AAB08926.
XX
```

```
PT   Human secreted proteins and coding sequences useful in diagnostic and
PT   therapeutic methods for disorders such as immune system or
PT   proliferative disorders, related to the proteins -
XX
PS   Claim 1; Page 347-348; 416pp; English.
XX
CC   The polynucleotide sequences given in AAA39052 to AAA39088 encode the
CC   human secreted proteins given in AAB08891 to AAB08984. The human secreted
CC   proteins can have activities based on the tissues and cells they are
CC   expressed in. Examples of the activities are: cytostatic;
CC   anti-proliferative; immunosuppressive; antibacterial; and vulnerary. The
CC   secreted proteins and their related polynucleotide sequences are useful
CC   for diagnostic and therapeutic methods useful for diagnosing and treating
CC   disorders related to the secreted proteins. The proteins, and
CC   polynucleotide sequences may be useful for treating disorders of the
CC   immune system, hyperproliferative disorders, infectious disease,
CC   regeneration of tissues, for chemotaxis and for screening molecules that
CC   bind to the proteins. The proteins or polynucleotide sequences may be
CC   used as food additives or preservatives, to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, co-factors or other nutritional components. Agonists or
CC   antagonists of the proteins may be used to prevent scar tissue growth
CC   during wound healing, and hyper-vascular diseases. AAA39043 to AAA39051
CC   and AAB08890 are sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1135 BP; 243 A; 333 C; 323 G; 234 T; 2 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 1135;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagc 19
          ||||||||||||||
Db    188 AGGTGCAGGTCAGC 175


RESULT  29
AAA91019/c
ID   AAA91019 standard; DNA; 1161 BP.
XX
AC   AAA91019;
XX
DT   05-APR-2001  (first entry)
XX
DE   Human secreted protein PRO7154 coding sequence.
XX
KW   Secreted protein; human; PRO protein; neoplastic cell growth; tumour;
KW   proliferation; leukaemia; lymphoid malignancy; inflammatory disorder;
KW   angiogenic disorder; immunologic disorder; PRO7154; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             102..1085
FT                   /*tag= a
FT                   /product= PRO7154
XX
PN   WO200075317-A2.
XX
PD   14-DEC-2000.
XX
PF   15-MAY-2000; 2000WO-US13358.
XX
PR   09-JUN-1999;   99US-0138385.
PR   20-JUL-1999;   99US-0144790.
PR   03-AUG-1999;   99US-0146843.
PR   10-AUG-1999;   99US-0148188.
PR   17-AUG-1999;   99US-0149320.
PR   17-AUG-1999;   99US-0149327.
PR   17-AUG-1999;   99US-0149396.
PR   20-AUG-1999;   99US-0150114.
```

PR 31-AUG-1999; 99US-0151700.
PR 31-AUG-1999; 99US-0151734.
XX
PA (GETH) GENENTECH INC.
XX
PI Botstein DA, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR WPI; 2001-071075/08.
DR P-PSDB; AAY97585.
XX
PT Antibodies against PRO polypeptides, useful for diagnosing and treating
PT tumours are associated with gene amplification, neoplastic cell growth
PT and proliferation in mammals -
XX
PS Claim 50; Fig 11; 143pp; English.
XX
CC This sequence encodes a human PRO protein of the invention. The PRO
CC proteins are secreted proteins. Antagonists or antibodies of PRO
CC polypeptides are useful for diagnosing and treating tumours are
CC associated with gene amplification, neoplastic cell growth and
CC proliferation in mammals, and those conditions characterised by
CC overexpression and/or activation of the amplified genes. Such conditions
CC include benign or malignant tumours (e.g. renal, liver, kidney, bladder,
CC breast, gastric, ovarian, colorectal, prostate, pancreatic, lung, vulval,
CC thyroid, hepatic carcinomas, sarcomas, glioblastomas and various head and
CC neck tumours); leukaemias and lymphoid malignancies; neuronal, glial,
CC astrocytal, hypothalamic, and other glandular, macrophageal, epithelial,
CC stromal and blastocoelic disorders; and inflammatory, angiogenic and
CC immunologic disorders. These may further be used to qualitatively or
CC quantitatively detect the expression of proteins encoded by the
CC amplified genes, and in tumour diagnostics or prognostics. The PRO
CC polypeptide or its antagonist may be used for the preparation of a
CC medicament in the treatment of a condition, which is responsive to the
CC PRO polypeptide, its antagonist or anti-PRO antibody.
XX
SQ Sequence 1161 BP; 235 A; 356 C; 327 G; 242 T; 1 other;

Query Match 56.0%; Score 14; DB 22; Length 1161;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 240 AGGTGCAGGTCAGC 227

RESULT 30
AAZ98136/c
ID AAZ98136 standard; cDNA; 1173 BP.
XX
AC AAZ98136;
XX
DT 11-MAY-2000 (first entry)
XX
DE Human signal peptide containing protein HSPP-28 cDNA SEQ ID NO:162.
XX
KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's diseases; ovulatory defect;
KW muscular dystrophy; ss.
XX
OS Homo sapiens.
XX
PN WO200000610-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14484.
XX
PR 26-JUN-1998; 98US-0090762.
PR 31-JUL-1998; 98US-0094983.
PR 01-OCT-1998; 98US-0102686.
PR 11-DEC-1998; 98US-0112129.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
PI Bandman O;
XX
DR WPI; 2000-160673/14.
DR P-PSDB; AAY87251.
XX
PT New human signal peptide-containing proteins useful in treatment,
PT prevention and diagnosis of e.g. cancer, inflammation and
PT cardiovascular disease -
XX
PS Claim 9; Page 268; 327pp; English.
XX
CC AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
CC anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
CC neuroprotective, cardiovascular and antiasthmatic activities, and can
CC be used in gene therapy. HSPPs can be used to treat or prevent disorders
CC associated with decreased activity or function of HSPP. Antagonists of
CC HSPP are used to treat or prevent disorders associated with increased
CC activity or function of HSPP. Such diseases include cell proliferation
CC (including cancer), inflammation, cardiovascular, neurological,
CC reproductive or developmental disorders, (e.g. arteriosclerosis,
CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC asthma, Crohn's disease, microbial or other infections, congestive or
CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC nucleic acids can be used for the recombinant production of HSPP, for
CC detecting HSPP in standard hybridisation and amplification assays (for
CC diagnosis and monitoring), in gene therapy, as antisense,
CC triplex-forming or ribozyme therapeutics, for detecting related sequences
CC or genetic variations, and for chromosomal mapping. HSPP are also used to
CC raise specific antibodies (Ab) and to screen for agonists and
CC antagonists (potential therapeutic agents). Ab are used to diagnose, or
CC monitor, HSPP-related diseases (in usual immunoassays), as therapeutic
CC antagonists, in competitive drug screens, and for purification of HSPP
CC from natural sources.
XX
SQ Sequence 1173 BP; 236 A; 363 C; 330 G; 244 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1173;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 256 AGGTGCAGGTCAGC 243

RESULT 31
AAC98064/c
ID AAC98064 standard; cDNA; 1186 BP.
XX
AC AAC98064;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human colon cancer antigen nucleotide sequence SEQ ID NO:74.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;

```
KW   neural disorder; immune system disorder; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; renal disorder;
KW   infectious disease; cardiovascular disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055351-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05883.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587534/55.
DR   P-PSDB; AAB53307.
XX
PT   Colon cancer associated gene sequences, referred to as colon cancer
PT   antigens, useful for the treatment, prevention, and diagnosis of colon
PT   disorders such as colon cancer -
XX
PS   Claim 1; Page 520-521; 2104pp; English.
XX
CC   AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC   called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC   human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC   neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC   vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC   can be used in gene therapy. The colon cancer antigen polynucleotides,
CC   proteins and antibodies to the proteins are useful for the prevention,
CC   treatment and diagnosis of colon disorders, such as colon cancer. The
CC   polynucleotides may be used in diagnostics and research, such as for
CC   chromosome identification, and as hybridisation probes. The proteins
CC   may also be used to prevent diseases such as neural disorders, immune
CC   system disorders, muscular disorders, reproductive disorders,
CC   gastrointestinal disorders, wounds, renal disorders, infectious
CC   diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC   AAB54007 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1186 BP; 244 A; 361 C; 338 G; 242 T; 1 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 1186;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     255 AGGTGCAGGTCAGC 242


RESULT  32
AAH35018/c
ID   AAH35018 standard; cDNA; 1186 BP.
XX
AC   AAH35018;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:2100.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; chromosome 11; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG75613.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 1; Page 3586; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence 1186 BP; 244 A; 362 C; 338 G; 242 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 1186;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     255 AGGTGCAGGTCAGC 242


RESULT  33
AAT98726
ID   AAT98726 standard; DNA; 1191 BP.
XX
AC   AAT98726;
XX
DT   09-NOV-1998  (first entry)
XX
DE   DNA encoding a S. pneumoniae protein of unknown function.
XX
KW   Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KW   immunological response; inoculation; antibody production; inhibitor;
KW   T cell immune response; antimicrobial compound; bacterial adhesion;
KW   extracellular matrix protein; protein-mediated cell invasion; wound;
KW   pathogenesis; ss.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             596..979
FT                   /*tag=  a
XX
```

PN WO9743303-A1.
XX
PD 20-NOV-1997.
XX
PF 14-MAY-1997; 97WO-US07950.
XX
PR 14-MAY-1996; 96US-0017670.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI Stodola RK;
XX
DR WPI; 1998-008793/01.
DR P-PSDB; AAW38682.
XX
PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial
PT infections
XX
PS Claim 4; Pages 222-223; 483pp; English.
XX
CC This sequence encodes a Streptococcus pneumoniae protein of unknown
CC function, and represents a DNA sequence of the invention.
CC The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and
CC inhibit or activate the activity of the proteins. Antagonists can be
CC used to treat diseases caused by S. pneumoniae proteins, through genetic
CC immunisation. They can also be used to induce an immunological response
CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
CC of the encoding nucleic acids in a vector adequate to produce antibody
CC and/or T cell immune responses to protect the animal from disease. The
CC proteins can also be used to identify antimicrobial compounds which are
CC capable of inhibiting their bioactivity. In particular the proteins of
CC the invention can be used to prevent adhesion of bacteria to mammalian
CC extracellular matrix proteins on in-dwelling devices or in wounds, to
CC block protein-mediated mammalian cell invasion, and to block the normal
CC progression of pathogenesis in infections initiated other than by the
CC implantation of in-dwelling devices or other surgical techniques.
XX
SQ Sequence 1191 BP; 294 A; 258 C; 294 G; 345 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 1191;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 cgtcaggtgcaggt 15
||||||||||||||
Db 954 cgtcaggtgcaggt 967

RESULT 34
AAF44623/c
ID AAF44623 standard; cDNA; 1368 BP.
XX
AC AAF44623;
XX
DT 27-MAR-2001 (first entry)
XX
DE Novel protein kinase cDNA, SEQ ID NO: 2.
XX
KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
KW immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
KW dermatological; antidiabetic; antiinfertility; gene therapy; vaccine;
KW immune disorder; cardiovascular disease; neurodegenerative disease;
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.
XX
OS Mus musculus.
XX

PN WO200073469-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14842.
XX
PR 28-MAY-1999; 99US-0136503.
XX
PA (SUGE-) SUGEN INC.
XX
PI Plowman GD, Martinez R, Whyte D, Sudersanam S;
XX
DR WPI; 2001-032161/04.
DR P-PSDB; AAB65598.
XX
PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and
PT treating immune-related diseases and disorders, cardiovascular disease,
PT neurodegenerative diseases and/or cancers -
XX
PS Disclosure; Fig 2; 310pp; English.
XX
CC The present sequence encodes a novel protein kinase. The nucleic acids
CC and the protein kinases they encode may be used in the treatment and
CC diagnosis of diseases associated with inappropriate kinase expression
CC such as immune-related diseases and disorders, cardiovascular disease,
CC neurodegenerative diseases and/or cancers. The nucleic acids and
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays. The kinase polypeptides may be used as antigens in the production
CC of antibodies of kinase expression and activity. Anti-kinase antibodies
CC and kinase antagonists may also be used to down regulate kinase
CC expression and activity. Diseases related to kinase expression and
CC activity include rheumatoid arthritis, atherosclerosis, autoimmune
CC disorders, complications of organ transplantation, myocardial infarction,
CC immune disorders, cardiomyopathies, strokes, renal failure,
CC oxidative-stress related disorders, chronic inflammatory bowel disease,
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC reproductive disorders.
XX
SQ Sequence 1368 BP; 346 A; 359 C; 398 G; 265 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1368;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 1022 AGGTGCAGGTCAGC 1009

RESULT 35
AAZ96242
ID AAZ96242 standard; DNA; 1634 BP.
XX
AC AAZ96242;
XX
DT 10-APR-2000 (first entry)
XX
DE S. pneumoniae derived DNA from ORF #70.
XX
KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
OS Streptococcus pneumoniae.
XX
PN WO9806734-A1.
XX
PD 19-FEB-1998.
XX
PF 15-AUG-1997; 97WO-US14436.
XX
PR 16-AUG-1996; 96US-0024022.

```
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Lonetto MA,  Nicholas RO;
PI   Stodola RK;
XX
DR   WPI; 1998-159452/14.
DR   P-PSDB; AAY85867.
XX
PT   Streptococcus pneumoniae proteins and related DNA - useful for
PT   screening compounds for antibacterial activity
XX
PS   Claim 4; Page 101-102; 640pp; English.
XX
CC   This invention describes novel isolated Streptococcus pneumoniae
CC   polynucleotides (see AAZ96173-Z96494) and their encoded proteins (see
CC   AAY85792-Y86182). The DNA, vectors and host cells described in the
CC   method of the invention are useful for the recombinant expression of the
CC   polypeptides. The polypeptides are useful for treatment or prevention of
CC   disease, or diagnosis of disease related to expression or activity of
CC   such a polypeptide. They can also be used to screen for compounds which
CC   interact with and inhibit or activate such a polypeptide. The
CC   polypeptides (or DNA encoding them, via gene therapy) are also useful
CC   for inducing an immunological response in a mammal. The antagonists are
CC   useful to inhibit such bacterial polypeptides. The polypeptides are
CC   particularly useful to identify antimicrobial compounds and antibiotics.
CC   They are also useful to determine their role in pathogenesis of
CC   infection, dysfunction and disease.
XX
SQ   Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;

  Query Match             56.0%;  Score 14;  DB 19;  Length 1634;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggt 15
           ||||||||||||||
Db     954 cgtcaggtgcaggt 967


RESULT  36
AAV37360
ID   AAV37360 standard; DNA; 1634 BP.
XX
AC   AAV37360;
XX
DT   13-OCT-1998  (first entry)
XX
DE   Streptococcus pneumoniae coding region.
XX
KW   coding region; ORF; open reading frame; antibacterial;
KW   infection; prevention; meningitis; ss.
XX
OS   Streptococcus pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1040..1291
FT                   /*tag=  a
FT                   /product= unknown
XX
PN   WO9819689-A1.
XX
PD   14-MAY-1998.
XX
PF   27-OCT-1997;   97WO-US19226.
XX
PR   01-NOV-1996;   96US-0029930.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Black MT,  Hodgson JE,  Knowles DJC,  Lonetto MA,  Nicholas RO;
PI   Reid RH,  Zarfos PN;
XX
DR   WPI; 1998-286586/25.
DR   P-PSDB; AAW60963.
XX
PT   New isolated nucleic acids from Streptococcus pneumoniae - useful,
PT   e.g. for identifying anti-bacterial(s) for treatment and prevention
PT   of meningitis
XX
PS   Claim 1; Page 60; 130pp; English.
XX
CC   The sequence is that of a coding region isolated from
CC   S. pneumoniae. Its encoded protein, or agonists of it,
CC   may be useful as an antibacterial for treatment or
CC   prevention of infection, specifically caused by S. pneumoniae
CC   (particularly meningitis) but possibly also Helicobacter
CC   pylori (ulcers and gastric cancer). It may be of particular
CC   use before insertion of an in-dwelling device or any other
CC   invasive procedure. The protein, or nucleic acid encoding
CC   it, can also be used in vaccines to induce a cellular
CC   and/or humoral immune response, or to screen for other
CC   antibacterials. The DNA may also contain flanking sequences
CC   that are potential sources of control elements for bacterial
CC   gene expression. Detecting a sequence encoding the protein
CC   can be used diagnostically, e.g. to detect a mutation for
CC   serotyping or classifying infectious agents.
XX
SQ   Sequence 1634 BP; 430 A; 339 C; 385 G; 480 T; 0 other;

  Query Match             56.0%;  Score 14;  DB 19;  Length 1634;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggt 15
           ||||||||||||||
Db     954 cgtcaggtgcaggt 967


RESULT  37
AAI59259/c
ID   AAI59259 standard; cDNA; 1814 BP.
XX
AC   AAI59259;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 1462.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
```

PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM40103.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 1462; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1814 BP; 447 A; 531 C; 481 G; 355 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1814;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 213 AGGTGCAGGTCAGC 200

RESULT 38
AAI61045
ID AAI61045 standard; cDNA; 1816 BP.
XX
AC AAI61045;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5034.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41889.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 5034; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1816 BP; 355 A; 481 C; 531 G; 449 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1816;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 aggtgcaggtcagc 19
||||||||||||||
Db 1604 aggtgcaggtcagc 1617

RESULT 39
AAA11992
ID AAA11992 standard; DNA; 37856 BP.
XX
AC AAA11992;
XX
DT 07-AUG-2000 (first entry)
XX
DE S. cellulosum DNA encoding polyketide and hereropolyketide enzymes.
XX
KW Polyketide biosynthesis; heteropolyketide biosynthesis; mutasynthesis;
KW epothilone; cytotoxic; immunosuppressant; antibiotic; antifungal;
KW plant-protection; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS complement (3398..6100)
FT /*tag= a
FT /product= "ORF1-tRNA synthetase"
FT /note= "gtg start codon"
FT CDS complement (6374..7111)
FT /*tag= b
FT /product= "ORF2-monooxygenase"
FT CDS complement (8433..9550)
FT /*tag= c

```
FT                   /product= "ORF3-aminotransferase"
FT                   /note= "AGT start codon given in the specification"
FT   CDS             9855..11393
FT                   /*tag=  d
FT                   /product= "ORF4- tyrosine/DOPA-Decarboxylase"
FT                   /note= "GTG start codon"
FT   CDS             12212..13658
FT                   /*tag=  e
FT                   /product= "ORF5-3-oxoacyl-ACP-reductase"
FT                   /note= "ACC start codon"
FT   CDS             15374..19984
FT                   /*tag=  f
FT                   /product= "ORF6-polyketide synthase"
FT   CDS             20003..27889
FT                   /*tag=  g
FT                   /product= "ORF7-peptide synthetase"
FT   CDS             28251..29400
FT                   /*tag=  h
FT                   /product= "ORF8-transpeptidase"
FT   CDS             complement (30040..31720)
FT                   /*tag=  i
FT                   /product= "ORF9-regulation element"
FT                   /note= "CGC stop codon"
FT   CDS             31982..32932
FT                   /*tag=  j
FT                   /product= "ORF10-transcription regulator"
FT   CDS             33128..33613
FT                   /*tag=  k
FT                   /product= "ORF11-regulation element"
FT                   /note= "GTG start codon"
FT   CDS             33661..34077
FT                   /*tag=  l
FT                   /product= "ORF12-regulation element"
FT   CDS             complement (35255..35616)
FT                   /*tag=  m
FT                   /product= "ORF13-transcription regulator"
FT   CDS             complement (35730..36242)
FT                   /*tag=  n
FT                   /product= "ORF14-transcription regulator"
FT                   /note= "GTG start codon"
XX
PN   DE19846493-A1.
XX
PD   13-APR-2000.
XX
PF   09-OCT-1998;   98DE-1046493.
XX
PR   09-OCT-1998;   98DE-1046493.
XX
PA   (GBFB ) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
PI   Beyer S,  Mueller R;
XX
DR   WPI; 2000-294101/26.
XX
PT   DNA sequence coding for products involved in the biosynthesis of
PT   polyketide or heteropolyketide compounds, especially epothilone  -
XX
PS   Claim 3; Page 20-33; 36pp; German.
XX
CC   This invention describes a novel DNA sequence (I) whose expression
CC   products effect or are involved in the enzymatic biosynthesis,
CC   mutasynthesis or partial synthesis of polyketide or heteropolyketide
CC   compounds (II). (I) can be inserted into an expression vector and used
CC   to transform or transfect prokaryotic or eukaryotic cells with the aim
CC   of obtaining strains that produce large amounts of polyketide or
CC   heteropolyketide compounds, especially epothilones, which have cytotoxic
CC   and/or immunosuppressant and antibiotic and antifungal activities and
CC   are useful as plant-protection agents. This sequence represents the DNA
CC   sequence isolated from Sorangium cellulosum which is described in the
CC   method of the invention.
XX
SQ   Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 other;
```

```
  Query Match             56.0%;  Score 14;  DB 21;  Length 37856;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtc 16
           ||||||||||||||
Db   35785 gtcaggtgcaggtc 35798


RESULT  40
AAV30458
ID   AAV30458 standard; DNA; 534720 BP.
XX
AC   AAV30458;
XX
DT   14-OCT-1998  (first entry)
XX
DE   Rhizobium species plasmid pNGR234a.
XX
KW   Symbiosis; open reading frame; ORF; plasmid; vector; transportation;
KW   degradation; metabolism; host range; nitrogen fixation; nodulation;
KW   legume; plant; ds.
XX
OS   Rhizobium sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             417796..418671
FT                   /*tag=  a
FT                   /standard_name= "ORF K1"
FT                   /product= "oligopeptide permease"
FT                   /note= "homologous to the OppC gene"
FT   CDS             418673..419680
FT                   /*tag=  b
FT                   /standard_name= "ORF K2"
FT                   /product= "oligopeptide permease"
FT                   /note= "homologous to the OppD gene"
FT   CDS             419677..420738
FT                   /*tag=  c
FT                   /standard_name= "ORF K3"
FT                   /product= "oligopeptide permease"
FT                   /note= "homologous to the OppF gene"
FT   CDS             420774..422159
FT                   /*tag=  d
FT                   /standard_name= "ORF K4"
FT                   /product= "encapsulation-like protein"
FT                   /note= "homologous to the CapA gene"
FT   CDS             422628..424031
FT                   /*tag=  e
FT                   /standard_name= "ORF K5"
FT                   /product= "aminotransferase-like protein"
FT                   /note= "homologous to the BioA gene"
FT   CDS             424056..425594
FT                   /*tag=  f
FT                   /standard_name= "ORF K6"
FT                   /product= "(semi)aldehyde dehydrogenase-like protein"
FT   CDS             complement (426949..428028)
FT                   /*tag=  g
FT                   /standard_name= "ORF K7"
FT                   /product= "transposase homologue"
FT                   /note= "homologous to the Tnp gene"
FT   CDS             428292..429623
FT                   /*tag=  h
FT                   /standard_name= "ORF K8"
FT                   /product= "glutamate dehydrogenase-like protein"
FT                   /note= "homologous to the GLUD1 gene"
FT   CDS             complement (430538..431284)
FT                   /*tag=  i
FT                   /standard_name= "ORF K9"
FT                   /product= "transposase homologue"
FT   CDS             complement (431296..432840)
FT                   /*tag=  j
```

```
FT                   /standard_name= "ORF K10"
FT                   /product= "transposase homologue"
FT                   /note= "homologous to the Tnp gene"
FT   CDS             complement (433880..434110)
FT                   /*tag= k
FT                   /standard_name= "ORF K11"
FT                   /product= "protein of unknown function"
FT                   /note= "homologous to the FixU gene"
FT   CDS             complement (434107..434433)
FT                   /*tag= l
FT                   /standard_name= "ORF K12"
FT                   /product= "protein of unknown function"
FT   CDS             complement (434517..434711)
FT                   /*tag= m
FT                   /standard_name= "ORF K13"
FT                   /product= "ferrodoxin/ferrodoxin-like protein"
FT                   /note= "homologous to the FdxN gene"
FT   CDS             complement (434753..436234)
FT                   /*tag= n
FT                   /standard_name= "ORF K14"
FT                   /gene= "nifB"
FT                   /product= "protein involved in FeMo co-factor
FT                            biosynthesis"
FT   CDS             complement (436460..438130)
FT                   /*tag= o
FT                   /standard_name= "ORF K15"
FT                   /gene= "nifA"
FT                   /product= "positive regulator of nif, fix and other
FT                            genes"
FT   CDS             complement (438297..438590)
FT                   /*tag= p
FT                   /standard_name= "ORF K16"
FT                   /gene= "fixX"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (438605..439912)
FT                   /*tag= q
FT                   /standard_name= "ORF K17"
FT                   /gene= "fixC"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (439923..441032)
FT                   /*tag= r
FT                   /standard_name= "ORF K18"
FT                   /gene= "fixB"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (441042..441899)
FT                   /*tag= s
FT                   /standard_name= "ORF K19"
FT                   /gene= "fixA"
FT                   /product= "protein required for nitrogenase activity"
FT   CDS             complement (442316..442636)
FT                   /*tag= t
FT                   /standard_name= "ORF K20
FT                   /product= "protein of unknown function"
FT   CDS             complement (443313..443879)
FT                   /*tag= u
FT                   /standard_name= "ORF K21"
FT                   /product= "protein of unknown function"
FT   CDS             444337..445029
FT                   /*tag= v
FT                   /standard_name= "ORF K22"
FT                   /product= "ferrodoxin-like protein"
FT                   /note= "homologous to the NifQ gene"
FT   CDS             445088..446602
FT                   /*tag= w
FT                   /standard_name= "ORF K23"
FT                   /gene= "dctA"
FT                   /product= "C4-dicarboxylate transport protein"
FT                   /note= "homologous to the DctAI gene"
FT   CDS             446599..447843
FT                   /*tag= x
FT                   /standard_name= "ORF L1"
FT                   /product= "cytochrome P450-like protein"
FT                   /note= "homologous to the CamC gene"

FT   CDS             447844..448500
FT                   /*tag= y
FT                   /standard_name= "ORF L2"
FT                   /product= "gamma-hexachlorocyclohexane-dechlorinase-like
FT                            protein"
FT                   /note= "homologous to the LinA gene"
FT   CDS             448497..450203
FT                   /*tag= z
FT                   /standard_name= "ORF L3"
FT                   /product= "putative protein with degradative function"
FT   CDS             450341..451396
FT                   /*tag= aa
FT                   /standard_name= "ORF L4"
FT                   /product= "luciferase alpha-subunit-like protein"
FT                   /note= "homologous to the LuxA gene"
FT   CDS             452980..454494
FT                   /*tag= ab
FT                   /standard_name= "ORF L6"
FT                   /gene= "nifD"
FT                   /product= "alpha-subunit of FeMo protein of nitrogenase"
FT   CDS             454590..456131
FT                   /*tag= ac
FT                   /standard_name= "ORF L7"
FT                   /gene= "nifK"
FT                   /product= "beta-subunit of FeMo protein of nitrogenase"
FT   CDS             456187..457677
FT                   /*tag= ad
FT                   /standard_name= "ORF L8"
FT                   /product= "protein involved in FeMo co-factor
FT                            biosynthesis"
FT                   /note= "homologous to the NifB gene"
FT   CDS             457687..459096
FT                   /*tag= ae
FT                   /standard_name= "ORF L9"
FT                   /product= "protein involved in FeMo co-factor
FT                            biosynthesis"
FT                   /note= "homologous to the FixF gene"
FT   CDS             459093..459575
FT                   /*tag= af
FT                   /standard_name= "ORF L10"
FT                   /product= "protein of unknown function"
FT                   /note= "homologous to the NifX gene"
FT   CDS             459579..460067
FT                   /*tag= ag
FT                   /standard_name= "ORF L11"
FT                   /product= "protein of unknown function"
FT   CDS             460501..460920
FT                   /*tag= ah
FT                   /standard_name= "ORF L12"
FT                   /product= "protein similar to part of the Fe protein
FT                            of nitrogenase"
FT                   /note= "homologous to the NifH gene"
FT   CDS             461228..461545
FT                   /*tag= ai
FT                   /standard_name= "ORF L13"
FT                   /product= "protein of unknown function"
FT   CDS             463201..464739
FT                   /*tag= aj
FT                   /standard_name= "ORF L14"
FT                   /product= "peptidase-like protein"
FT                   /note= "homologous to the bI-MPP gene"
FT   CDS             464736..466079
FT                   /*tag= ak
FT                   /standard_name= "ORF L15"
FT                   /product= "processing protease-like protein"
FT                   /note= "homologous to the PP gene"
FT   CDS             466590..467021
```

```
  Query Match           56.0%;  Score 14;  DB 19;  Length 534720;
  Best Local Similarity  100.0%;  Pred. No. 89;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtca 17
```

```
          ||||||||||||||
Db 435493 tcaggtgcaggtca 435506


RESULT  41
AAV30459
ID   AAV30459 standard; DNA; 536165 BP.
XX
AC   AAV30459;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Rhizobium species symbiotic plasmid pNGR234.
XX
KW   Symbiosis; open reading frame; ORF; plasmid; vector; transportation;
KW   degradation; metabolism; host range; nitrogen fixation; nodulation;
KW   legume; plant; ds.
XX
OS   Rhizobium sp.
XX
PN   WO9802560-A2.
XX
PD   22-JAN-1998.
XX
PF   10-JUL-1997;   97WO-IB00950.
XX
PR   20-MAY-1997;   97GB-0010395.
PR   12-JUL-1996;   96EP-0730001.
XX
PA   (MOLE-) INST MOLECULAR BIOTECHNOLOGY.
PA   (BIOL-) LAB BIOLOGIE MOLECULAIRE PLANTES SUPERIE.
XX
PI   Broughton WJ,  Freiberg CB,  Perret XP,  Rosenthal A;
XX
DR   WPI; 1998-110606/10.
XX
PT   New isolated symbiotic plasmid from Rhizobium sp. NGR234 - used to
PT   develop products for modifying plant characteristics, e.g. nitrogen
PT   fixation, synthesis of compounds and stress response
XX
PS   Claim 1; Fig 3; 228pp; English.
XX
CC   This is the nucleotide sequence of the plasmid pNGR234a isolated from
CC   Rhizobium sp. NGR234.  Open reading frames (ORF) derivable from the
CC   nucleotide sequence are claimed.  The nucleotide sequences or ORFs can
CC   be used e.g. in the transportation of compounds to and from an organism
CC   which is a host to at least one of the nucleotide sequences, ORFs or
CC   proteins, the degradation and/or metabolism of organic, inorganic,
CC   natural or xenobiotic substances in a host organism or the modification
CC   of the host range, nitrogen fixation abilities; for obtaining a synthetic
CC   minimal set of ORFs required for functional Rhizobium-legume symbiosis,
CC   especially for nodulation efficiency on host plants.
XX
SQ   Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 19;  Length 536165;
  Best Local Similarity   100.0%;  Pred. No. 89;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtca 17
           ||||||||||||||
Db 435493 tcaggtgcaggtca 435506


RESULT  42
AAQ13647
ID   AAQ13647 standard; DNA; 18 BP.
XX
AC   AAQ13647;
XX
DT   28-NOV-1991  (first entry)
XX
DE   Exon B/intron junction of PTPase B1 gene.
XX
KW   PTPB1; protein phosphotyrosyl phosphatase B1;
KW   growth suppression activity; ANLL; myeloproliferation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            1..9
FT                   /*tag=  a
FT                   /note= "last 3 codons of exon B"
FT   intron          10..18
FT                   /*tag=  b
FT                   /note= "5' end of intron"
XX
PN   WO9113173-A.
XX
PD   05-SEP-1991.
XX
PF   01-MAR-1991;   91WO-0001432.
XX
PR   18-JAN-1991;   91US-0643041.
PR   02-MAR-1990;   90US-0487733.
XX
PA   (BIOT-) APPL BIOTECHN INC.
XX
PI   Bruskin AM,  Hill DE;
XX
DR   WPI; 1991-281485/38.
XX
PT   Vector contg. protein phospho:tyrosyl phosphatase gene fragment -
PT   for use in diagnosis of cancers, e.g. acute non-lymphocytic
PT   leukaemia
XX
PS   Disclosure; Fig 4B; 73pp; English.
XX
CC   The PTPase B1 genomic sequence comprises a number of exons and
CC   introns. Exon B encodes amino acids Met(235) to Gln(288) of the
CC   PTPase B1 protein (see AAR13907). This is the junction of the 3' end
CC   of exon B and the 5' end of the intron.
CC   See AAQ13643-Q13654.
XX
SQ   Sequence 18 BP; 3 A; 5 C; 5 G; 5 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 12;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 4.7e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 gtgcaggtcagca 20
           |||||||||||||
Db       4 gtgcaggtcagca 16


RESULT  43
AAX85925/c
ID   AAX85925 standard; DNA; 123 BP.
XX
AC   AAX85925;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Oligonucleotide used to produce heavy chain variable region of Ig NEW.
XX
KW   Light chain variable region; interleukin-4; IL-4; antibody 3B9;
KW   chimeric antibody; humanised antibody; IL-4 mediated allergic reaction;
KW   Immunoglobulin E-mediated allergic reaction; allergic rhinitis;
KW   conjunctivitis; atopic dermatitis; atopic asthma; anaphylactic shock;
KW   rheumatoid arthritis; host-versus-graft disease; renal disease;
KW   allergy; Ig NEW; ss.
XX
OS   Synthetic.
XX
```

```
PN   US5928904-A.
XX
PD   27-JUL-1999.
XX
PF   07-JUN-1995;   95US-0483632.
XX
PR   07-JUN-1995;   95US-0483632.
PR   07-SEP-1993;   93US-0117366.
PR   14-OCT-1993;   93US-0136783.
PR   07-SEP-1994;   94WO-US10308.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Gross MS,  Holmes SD,  Sylvester DR;
XX
DR   WPI; 1999-429500/36.
XX
PT   New DNA molecules encoding recombinant antibodies useful for
PT   treating IL4-mediated conditions
XX
PS   Disclosure; Columns 61-63; 50pp; English.
XX
CC   The specification describes chimeric and humanised IL-4
CC   monoclonal antibodies. The antibodies of the invention are used in
CC   therapeutic and pharmaceutical compositions for treating IL-4 mediated
CC   and Immunoglobulin E-mediated allergic reactions e.g. allergic rhinitis,
CC   conjunctivitis, atopic dermatitis, atopic asthma, anaphylactic shock,
CC   rheumatoid arthritis, host-versus-graft disease and renal disease.
CC   They are also useful in the diagnosis of an allergy or condition
CC   associated with excess IL-4 production through the measurement e.g. by
CC   ELISA of circulating endogenous IL-4 levels in humans. Oligonucleotides
CC   AAX85925-28 were used to produce the heavy chain variable region of
CC   Ig NEW, in the course of the invention.
XX
SQ   Sequence 123 BP; 25 A; 37 C; 41 G; 20 T; 0 other;


  Query Match             52.0%;  Score 13;  DB 20;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 4.3e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  44
AAI23772
ID   AAI23772 standard; DNA; 153 BP.
XX
AC   AAI23772;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #13705 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 13705; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 153 BP; 33 A; 61 C; 29 G; 29 T; 1 other;


  Query Match             52.0%;  Score 13;  DB 22;  Length 153;
  Best Local Similarity   100.0%;  Pred. No. 4.2e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 tgcaggtcagcac 21
           |||||||||||||
Db       4 tgcaggtcagcac 16


RESULT  45
AAI49080
ID   AAI49080 standard; DNA; 153 BP.
XX
AC   AAI49080;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #17766 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
```

PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 17766; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 153 BP; 33 A; 61 C; 29 G; 29 T; 1 other;

Query Match 52.0%; Score 13; DB 22; Length 153;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 tgcaggtcagcac 21
||||||||||||
Db 4 tgcaggtcagcac 16

Search completed: December 26, 2001, 14:17:19
Job time: 7031 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:55:41 ; Search time 143.1 Seconds
(without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 9

Total number of hits satisfying chosen parameters: 13726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
c	5	15	60.0	630	3	US-08-545-809A-10	Sequence 10, Appl
c	6	14	56.0	613	3	US-08-545-809A-5	Sequence 5, Appli
	7	14	56.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	8	14	56.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	9	14	56.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	10	13	52.0	123	2	US-08-483-636-59	Sequence 59, Appl
c	11	13	52.0	123	2	US-08-483-632-59	Sequence 59, Appl
c	12	13	52.0	363	2	US-08-483-636-63	Sequence 63, Appl
c	13	13	52.0	363	2	US-08-483-632-63	Sequence 63, Appl
c	14	13	52.0	378	1	US-08-488-376-13	Sequence 13, Appl
c	15	13	52.0	378	1	US-08-488-376-15	Sequence 15, Appl
c	16	13	52.0	378	2	US-08-634-223-13	Sequence 13, Appl
c	17	13	52.0	378	2	US-08-634-223-15	Sequence 15, Appl
c	18	13	52.0	378	2	US-08-634-224-13	Sequence 13, Appl
c	19	13	52.0	378	2	US-08-634-224-15	Sequence 15, Appl
c	20	13	52.0	378	2	US-08-634-400-13	Sequence 13, Appl
c	21	13	52.0	378	2	US-08-634-400-15	Sequence 15, Appl
c	22	13	52.0	378	2	US-08-635-878-13	Sequence 13, Appl
c	23	13	52.0	378	2	US-08-635-878-15	Sequence 15, Appl
c	24	13	52.0	378	2	US-08-770-057-13	Sequence 13, Appl
c	25	13	52.0	378	2	US-08-770-057-15	Sequence 15, Appl
c	26	13	52.0	378	4	US-09-335-697B-13	Sequence 13, Appl
c	27	13	52.0	378	4	US-09-335-697B-15	Sequence 15, Appl
c	28	13	52.0	423	2	US-08-345-321-9	Sequence 9, Appli
c	29	13	52.0	546	3	US-08-545-809A-26	Sequence 26, Appl
	30	13	52.0	718	3	US-09-154-083-31	Sequence 31, Appl
	31	13	52.0	750	4	US-09-183-959-18	Sequence 18, Appl
c	32	13	52.0	879	3	US-08-714-071-3	Sequence 3, Appli
	33	13	52.0	1040	4	US-09-183-959-11	Sequence 11, Appl
c	34	13	52.0	1173	3	US-08-706-216-5	Sequence 5, Appli
c	35	13	52.0	1428	1	US-08-488-376-17	Sequence 17, Appl
c	36	13	52.0	1428	1	US-08-488-376-19	Sequence 19, Appl
c	37	13	52.0	1428	2	US-08-634-223-17	Sequence 17, Appl
c	38	13	52.0	1428	2	US-08-634-223-19	Sequence 19, Appl
c	39	13	52.0	1428	2	US-08-634-224-17	Sequence 17, Appl
c	40	13	52.0	1428	2	US-08-634-224-19	Sequence 19, Appl
c	41	13	52.0	1428	2	US-08-634-400-17	Sequence 17, Appl
c	42	13	52.0	1428	2	US-08-634-400-19	Sequence 19, Appl
c	43	13	52.0	1428	2	US-08-635-878-17	Sequence 17, Appl
c	44	13	52.0	1428	2	US-08-635-878-19	Sequence 19, Appl
c	45	13	52.0	1428	2	US-08-770-057-17	Sequence 17, Appl
c	46	13	52.0	1428	2	US-08-770-057-19	Sequence 19, Appl
c	47	13	52.0	1428	4	US-09-335-697B-17	Sequence 17, Appl
c	48	13	52.0	1428	4	US-09-335-697B-19	Sequence 19, Appl
	49	13	52.0	1640	4	US-08-676-444-41	Sequence 41, Appl
	50	13	52.0	2830	2	US-09-010-928B-1	Sequence 1, Appli
	51	13	52.0	3105	4	US-08-542-635-1	Sequence 1, Appli
	52	13	52.0	4465	2	US-08-620-605D-1	Sequence 1, Appli
	53	13	52.0	4533	3	US-08-726-214-5	Sequence 5, Appli
	54	13	52.0	4547	2	US-09-005-232A-1	Sequence 1, Appli
c	55	13	52.0	15239	1	US-08-390-878-17	Sequence 17, Appl
	56	13	52.0	49272	1	US-08-614-770A-1	Sequence 1, Appli
c	57	12	48.0	19	1	US-07-936-421-11	Sequence 11, Appl
	58	12	48.0	20	3	US-08-545-809A-72	Sequence 72, Appl
c	59	12	48.0	28	2	US-08-353-372A-2	Sequence 2, Appli
c	60	12	48.0	40	2	US-08-805-918-11	Sequence 11, Appl
c	61	12	48.0	40	2	US-08-811-028-7	Sequence 7, Appli
c	62	12	48.0	59	2	US-08-116-778E-24	Sequence 24, Appl
c	63	12	48.0	59	2	US-08-438-562-24	Sequence 24, Appl
c	64	12	48.0	59	2	US-08-483-528B-24	Sequence 24, Appl
c	65	12	48.0	59	3	US-08-673-799C-24	Sequence 24, Appl
	66	12	48.0	64	1	US-08-388-672A-8	Sequence 8, Appli
	67	12	48.0	64	3	US-09-080-554-8	Sequence 8, Appli
	68	12	48.0	66	1	US-08-207-996-5	Sequence 5, Appli
	69	12	48.0	66	2	US-08-760-840A-5	Sequence 5, Appli
	70	12	48.0	66	3	US-09-266-119-5	Sequence 5, Appli
c	71	12	48.0	70	1	US-08-290-592E-39	Sequence 39, Appl
c	72	12	48.0	70	5	PCT-US96-09448-39	Sequence 39, Appl
	73	12	48.0	75	1	US-08-657-012-5	Sequence 5, Appli
	74	12	48.0	75	3	US-09-013-872-5	Sequence 5, Appli
	75	12	48.0	75	4	US-09-184-198-5	Sequence 5, Appli
c	76	12	48.0	76	4	US-09-025-769B-134	Sequence 134, App
	77	12	48.0	93	4	US-08-556-978B-67	Sequence 67, Appl
c	78	12	48.0	93	4	US-08-556-978B-68	Sequence 68, Appl
	79	12	48.0	138	1	US-08-467-420A-58	Sequence 58, Appl
	80	12	48.0	138	1	US-08-470-110A-58	Sequence 58, Appl
	81	12	48.0	138	1	US-08-667-769A-58	Sequence 58, Appl
	82	12	48.0	138	2	US-08-940-371-58	Sequence 58, Appl
	83	12	48.0	138	5	PCT-US95-17082A-58	Sequence 58, Appl
c	84	12	48.0	152	1	US-08-136-277-9	Sequence 9, Appli
c	85	12	48.0	152	2	US-08-479-403-9	Sequence 9, Appli
c	86	12	48.0	152	3	US-08-835-734-9	Sequence 9, Appli
c	87	12	48.0	162	1	US-08-275-916-1	Sequence 1, Appli
c	88	12	48.0	162	5	PCT-US93-03868-1	Sequence 1, Appli
c	89	12	48.0	198	2	US-08-414-657D-15	Sequence 15, Appl
c	90	12	48.0	198	2	US-08-414-657D-16	Sequence 16, Appl
c	91	12	48.0	201	2	US-08-222-719-8	Sequence 8, Appli
c	92	12	48.0	201	2	US-08-470-925-8	Sequence 8, Appli
c	93	12	48.0	201	2	US-08-471-613-8	Sequence 8, Appli
c	94	12	48.0	201	5	PCT-US93-10443-8	Sequence 8, Appli
	95	12	48.0	303	4	US-08-556-978B-82	Sequence 82, Appl
c	96	12	48.0	357	1	US-08-459-310-11	Sequence 11, Appl
c	97	12	48.0	357	1	US-08-467-420A-61	Sequence 61, Appl
c	98	12	48.0	357	1	US-08-470-110A-61	Sequence 61, Appl
c	99	12	48.0	357	1	US-08-667-769A-61	Sequence 61, Appl
c	100	12	48.0	357	2	US-08-940-371-61	Sequence 61, Appl

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     111 TCGTCAGGTGCAGGTCAGCACGTTG 135


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     111 TCGTCAGGTGCAGGTCAGCACGTTG 135


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORF1"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORF1 RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tcgtcaggtgcaggtcagcacgttg 25
           |||||||||||||||||||||||||
Db     227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtcagcacgttg 25
          |||||||||||||||||||||||||
Db    227 TCGTCAGGTGCAGGTCAGCACGTTG 251


RESULT   5
US-08-545-809A-10/c
; Sequence 10, Application US/08545809A
; Patent No. 6096878
;  GENERAL INFORMATION:
;    APPLICANT:  Honjo, Tasuku
;    APPLICANT:  Matsuda, Fumihiko
;    TITLE OF INVENTION:  HUMAN IMMUNOGLOBULIN VH GENE
;    TITLE OF INVENTION:  SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
;    NUMBER OF SEQUENCES:  145
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/545,809A
;      FILING DATE:  27-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/00603
;      FILING DATE:  10-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  06501/004001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-542-5070
;      TELEFAX:  617-542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  630 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  human lymphoblast
;      CELL LINE:  CGM1
US-08-545-809A-10


  Query Match             60.0%;  Score 15;  DB 3;  Length 630;
  Best Local Similarity   100.0%;  Pred. No. 9;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcagca 20
          ||||||||||||||||
Db    280 AGGTGCAGGTCAGCA 266


RESULT   6
US-08-545-809A-5/c
; Sequence 5, Application US/08545809A
; Patent No. 6096878
;  GENERAL INFORMATION:
;    APPLICANT:  Honjo, Tasuku
;    APPLICANT:  Matsuda, Fumihiko
;    TITLE OF INVENTION:  HUMAN IMMUNOGLOBULIN VH GENE
;    TITLE OF INVENTION:  SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
;    NUMBER OF SEQUENCES:  145
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  Windows95
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/545,809A
;       FILING DATE:  27-MAR-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/JP93/00603
;       FILING DATE:  10-MAY-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Freeman, John W.
;       REGISTRATION NUMBER:  29,066
;       REFERENCE/DOCKET NUMBER:  06501/004001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-542-5070
;       TELEFAX:  617-542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  613 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;       CELL TYPE:  human lymphoblast
;       CELL LINE:  CGM1
US-08-545-809A-5


  Query Match             56.0%;  Score 14;  DB 3;  Length 613;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagc 19
           ||||||||||||||
Db     283 AGGTGCAGGTCAGC 270


RESULT   7
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             56.0%;  Score 14;  DB 4;  Length 4403765;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 gtgcaggtcagcac 21
             ||||||||||||||
Db 586964 gtgcaggtcagcac 586977


RESULT   8
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2


  Query Match             56.0%;  Score 14;  DB 4;  Length 4403765;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagca 20
             ||||||||||||||
Db 2740489 GGTGCAGGTCAGCA 2740476


RESULT   9
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1


  Query Match             56.0%;  Score 14;  DB 4;  Length 4411529;
  Best Local Similarity   100.0%;  Pred. No. 15;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagca 20
             ||||||||||||||
Db 2743678 GGTGCAGGTCAGCA 2743665
```

```
RESULT  10
US-08-483-636-59/c
; Sequence 59, Application US/08483636
; Patent No. 5914110
;  GENERAL INFORMATION:
;    APPLICANT:  Holmes, Stephen D.
;    APPLICANT:  Gross, Mitchell S.
;    APPLICANT:  Sylvester, Daniel R.
;    TITLE OF INVENTION:  Recombinant IL4 Antibodies Useful in
;    TITLE OF INVENTION:  Treatment of IL4 Mediated Disorders
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corp./Corporate
;      ADDRESSEE:  Intellectual Property
;      STREET:  P.O. Box 1539 / UW2220
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,636
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/117366
;      FILING DATE:  07-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/136783
;      FILING DATE:  14-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US/94/10308
;      FILING DATE:  07-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sutton, Jeffrey A.
;      REGISTRATION NUMBER:  34,028
;      REFERENCE/DOCKET NUMBER:  P50186-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 270-5024
;      TELEFAX:  (215) 270-5090
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  123 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-483-636-59


  Query Match             52.0%;  Score 13;  DB 2;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  11
US-08-483-632-59/c
; Sequence 59, Application US/08483632
; Patent No. 5928904
;  GENERAL INFORMATION:
;    APPLICANT:  Holmes, Stephen D.
;    APPLICANT:  Gross, Mitchell S.
;    APPLICANT:  Sylvester, Daniel R.
;    TITLE OF INVENTION:  Recombinant IL4 Antibodies Useful in
;    TITLE OF INVENTION:  Treatment of IL4 Mediated Disorders
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corp./Corporate
;      ADDRESSEE:  Intellectual Property
;      STREET:  P.O. Box 1539 / UW2220
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,632
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/117366
;      FILING DATE:  07-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/136783
;      FILING DATE:  14-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US/94/10308
;      FILING DATE:  07-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sutton, Jeffrey A.
;      REGISTRATION NUMBER:  34,028
;      REFERENCE/DOCKET NUMBER:  P50186-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 270-5024
;      TELEFAX:  (215) 270-5090
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  123 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-483-632-59


  Query Match             52.0%;  Score 13;  DB 2;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  12
US-08-483-636-63/c
; Sequence 63, Application US/08483636
; Patent No. 5914110
;  GENERAL INFORMATION:
;    APPLICANT:  Holmes, Stephen D.
;    APPLICANT:  Gross, Mitchell S.
;    APPLICANT:  Sylvester, Daniel R.
;    TITLE OF INVENTION:  Recombinant IL4 Antibodies Useful in
;    TITLE OF INVENTION:  Treatment of IL4 Mediated Disorders
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corp./Corporate
;      ADDRESSEE:  Intellectual Property
;      STREET:  P.O. Box 1539 / UW2220
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,636
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/117366
;      FILING DATE:  07-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/136783
;      FILING DATE:  14-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US/94/10308
;      FILING DATE:  07-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sutton, Jeffrey A.
;      REGISTRATION NUMBER:  34,028
;      REFERENCE/DOCKET NUMBER:  P50186-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 270-5024
;      TELEFAX:  (215) 270-5090
;  INFORMATION FOR SEQ ID NO:  63:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-483-636-63


  Query Match             52.0%;  Score 13;  DB 2;  Length 363;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  13
US-08-483-632-63/c
; Sequence 63, Application US/08483632
; Patent No. 5928904
;  GENERAL INFORMATION:
;    APPLICANT:  Holmes, Stephen D.
;    APPLICANT:  Gross, Mitchell S.
;    APPLICANT:  Sylvester, Daniel R.
;    TITLE OF INVENTION:  Recombinant IL4 Antibodies Useful in
;    TITLE OF INVENTION:  Treatment of IL4 Mediated Disorders
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corp./Corporate
;      ADDRESSEE:  Intellectual Property
;      STREET:  P.O. Box 1539 / UW2220
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,632
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/117366
;      FILING DATE:  07-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/136783
;      FILING DATE:  14-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US/94/10308
;      FILING DATE:  07-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sutton, Jeffrey A.
;      REGISTRATION NUMBER:  34,028
;      REFERENCE/DOCKET NUMBER:  P50186-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 270-5024
;      TELEFAX:  (215) 270-5090
;  INFORMATION FOR SEQ ID NO:  63:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-483-632-63


  Query Match             52.0%;  Score 13;  DB 2;  Length 363;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  14
US-08-488-376-13/c
; Sequence 13, Application US/08488376
; Patent No. 5811524
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
```

```
;       LENGTH:  378 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..378
US-08-488-376-13


  Query Match             52.0%;  Score 13;  DB 1;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  15
US-08-488-376-15/c
; Sequence 15, Application US/08488376
; Patent No. 5811524
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-488-376-15


  Query Match             52.0%;  Score 13;  DB 1;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  16
US-08-634-223-13/c
; Sequence 13, Application US/08634223
; Patent No. 5840298
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,223
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-634-223-13


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  17
US-08-634-223-15/c
; Sequence 15, Application US/08634223
```

```
; Patent No. 5840298
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,223
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-634-223-15


  Query Match           52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity  100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db     70 AGGTGCAGGTCAG 58


RESULT  18
US-08-634-224-13/c
; Sequence 13, Application US/08634224
; Patent No. 5866125
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,224
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-634-224-13


  Query Match           52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity  100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagc 19
          |||||||||||||
Db     69 GGTGCAGGTCAGC 57


RESULT  19
US-08-634-224-15/c
; Sequence 15, Application US/08634224
; Patent No. 5866125
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/634,224
;       FILING DATE:
;       CLASSIFICATION:  424
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/488,376
;       FILING DATE:  07-JUN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Teskin, Robin L.
;       REGISTRATION NUMBER:  35,030
;       REFERENCE/DOCKET NUMBER:  012712-150
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6620
;       TELEFAX:  (703) 836-2021
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  378 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..378
US-08-634-224-15


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  20
US-08-634-400-13/c
; Sequence 13, Application US/08634400
; Patent No. 5939068
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,400
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-634-400-13


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  21
US-08-634-400-15/c
; Sequence 15, Application US/08634400
; Patent No. 5939068
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,400
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-634-400-15


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  22
US-08-635-878-13/c
; Sequence 13, Application US/08635878
; Patent No. 5955364
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/635,878
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-635-878-13


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  23
US-08-635-878-15/c
; Sequence 15, Application US/08635878
; Patent No. 5955364
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/635,878
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-635-878-15


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58
```

```
RESULT  24
US-08-770-057-13/c
; Sequence 13, Application US/08770057
; Patent No. 5958765
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,057
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-770-057-13


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  25
US-08-770-057-15/c
; Sequence 15, Application US/08770057
; Patent No. 5958765
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,057
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  378 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..378
US-08-770-057-15


  Query Match             52.0%;  Score 13;  DB 2;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  26
US-09-335-697B-13/c
; Sequence 13, Application US/09335697B
; Patent No. 6200804
;   GENERAL INFORMATION:
;        APPLICANT: BRAMS, Peter
;                   CHAMAT, Soulaima Salim
;                   PAN, Li-Zhen
;                   WALSH, Edward E.
;                   HEARD, Cheryl Janne
;                   NEWMAN, Roland Anthony
;        TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
;                            MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;                            METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREO
;        NUMBER OF SEQUENCES: 19
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
```

```
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/335,697B
;             FILING DATE: 06-Jul-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/770,057
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Teskin, Robin L.
;             REGISTRATION NUMBER: 35,030
;             REFERENCE/DOCKET NUMBER: 012712-150
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 13:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 378 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..378
;        SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-335-697B-13


  Query Match             52.0%;  Score 13;  DB 4;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db      69 GGTGCAGGTCAGC 57


RESULT  27
US-09-335-697B-15/c
; Sequence 15, Application US/09335697B
; Patent No. 6200804
;   GENERAL INFORMATION:
;        APPLICANT: BRAMS, Peter
;                   CHAMAT, Soulaima Salim
;                   PAN, Li-Zhen
;                   WALSH, Edward E.
;                   HEARD, Cheryl Janne
;                   NEWMAN, Roland Anthony
;        TITLE OF INVENTION: NEUTRALIZING HIGH AFFINITY HUMAN
;                            MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;                            METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;        NUMBER OF SEQUENCES: 19
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Burns, Doane, Swecker & Mathis
;             STREET: P.O. Box 1404
;             CITY: Alexandria
;             STATE: Virginia
;             COUNTRY: United States
;             ZIP: 22313-1404
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/335,697B
;             FILING DATE: 06-Jul-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/770,057
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Teskin, Robin L.
;             REGISTRATION NUMBER: 35,030
;             REFERENCE/DOCKET NUMBER: 012712-150
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (703) 836-6620
;             TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 15:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 378 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..378
;        SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-335-697B-15


  Query Match             52.0%;  Score 13;  DB 4;  Length 378;
  Best Local Similarity   100.0%;  Pred. No. 98;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db      70 AGGTGCAGGTCAG 58


RESULT  28
US-08-345-321-9/c
; Sequence 9, Application US/08345321
; Patent No. 5914109
;  GENERAL INFORMATION:
;    APPLICANT:  ZOLLA-PAZNER, Susan
;    APPLICANT:  GORNY, Miroslav K.
;    TITLE OF INVENTION:  HETEROHYBRIDOMAS PRODUCING HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES TO HIV-1
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Browdy and Neimark
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/345,321
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/872,675
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Browdy, Roger L.
;      REGISTRATION NUMBER:  25,618
;      REFERENCE/DOCKET NUMBER:  ZOLLA-PAZNER1B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  9:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..423
US-08-345-321-9


  Query Match            52.0%;  Score 13;  DB 2;  Length 423;
  Best Local Similarity  100.0%;  Pred. No. 97;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 aggtgcaggtcag 18
          |||||||||||||
Db    127 AGGTGCAGGTCAG 115


RESULT  29
US-08-545-809A-26/c
; Sequence 26, Application US/08545809A
; Patent No. 6096878
;  GENERAL INFORMATION:
;    APPLICANT:  Honjo, Tasuku
;    APPLICANT:  Matsuda, Fumihiko
;    TITLE OF INVENTION:  HUMAN IMMUNOGLOBULIN VH GENE
;    TITLE OF INVENTION:  SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
;    NUMBER OF SEQUENCES:  145
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson, P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows95
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/545,809A
;      FILING DATE:  27-MAR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/JP93/00603
;      FILING DATE:  10-MAY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Freeman, John W.
;      REGISTRATION NUMBER:  29,066
;      REFERENCE/DOCKET NUMBER:  06501/004001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-542-5070
;      TELEFAX:  617-542-8906
;      TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  546 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      CELL TYPE:  human lymphoblast
;      CELL LINE:  CGM1
US-08-545-809A-26


  Query Match            52.0%;  Score 13;  DB 3;  Length 546;
  Best Local Similarity  100.0%;  Pred. No. 96;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagc 19
          |||||||||||||
Db    232 GGTGCAGGTCAGC 220


RESULT  30
US-09-154-083-31
; Sequence 31, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
;  APPLICANT: Wu, Kai
;  TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
;  TITLE OF INVENTION: Constructs Therefor
;  FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
;  CURRENT APPLICATION NUMBER: US/09/154,083
;  CURRENT FILING DATE: 1998-09-16
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 718
;   TYPE: DNA
;   ORGANISM: Streptomyces hygroscopicus
US-09-154-083-31


  Query Match            52.0%;  Score 13;  DB 3;  Length 718;
  Best Local Similarity  100.0%;  Pred. No. 94;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 gcaggtcagcacg 22
          |||||||||||||
Db    347 gcaggtcagcacg 359


RESULT  31
US-09-183-959-18
; Sequence 18, Application US/09183959
; Patent No. 6303332
; GENERAL INFORMATION:
;  APPLICANT: Cahoon, Rebecca E.
;  APPLICANT: Miao, Guo-Hua
;  APPLICANT: Rafalski, J. Antoni
;  APPLICANT: Taramino, Graziana
;  TITLE OF INVENTION: CORN cDNA ENCODING SOUTHERN LEAF BLIGHT RESISTANCE
;  FILE REFERENCE: BB-1125
;  CURRENT APPLICATION NUMBER: US/09/183,959
;  CURRENT FILING DATE: 1998-11-02
;  EARLIER APPLICATION NUMBER: 60/064,493
;  EARLIER FILING DATE: No. 6303332ember 5, 1997
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: Microsoft Windows 95
; SEQ ID NO 18
;   LENGTH: 750
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-183-959-18


  Query Match            52.0%;  Score 13;  DB 4;  Length 750;
  Best Local Similarity  100.0%;  Pred. No. 94;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagc 19
          |||||||||||||
Db    687 ggtgcaggtcagc 699


RESULT  32
US-08-714-071-3/c
; Sequence 3, Application US/08714071
; Patent No. 6136584
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Tsutomu, FUJIWARA
;    APPLICANT:  Shiro, OKUNO
;    APPLICANT:  Hisanobu, HIRANO
;    APPLICANT:  Sadahito, SHIN
;    TITLE OF INVENTION:  FK506 BINDING PROTEIN GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/714,071
;      FILING DATE:
;      CLASSIFICATION:  435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  879 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Human fetal brain cDNA library
;    IMMEDIATE SOURCE:
;      CLONE:  OTK4(6-1)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  70..393
US-08-714-071-3


  Query Match             52.0%;  Score 13;  DB 3;  Length 879;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     303 GGTGCAGGTCAGC 291


RESULT  33
US-09-183-959-11
; Sequence 11, Application US/09183959
; Patent No. 6303332
; GENERAL INFORMATION:
;  APPLICANT: Cahoon, Rebecca E.
;  APPLICANT: Miao, Guo-Hua
;  APPLICANT: Rafalski, J. Antoni
;  APPLICANT: Taramino, Graziana
;  TITLE OF INVENTION: CORN cDNA ENCODING SOUTHERN LEAF BLIGHT RESISTANCE
;  FILE REFERENCE: BB-1125
;  CURRENT APPLICATION NUMBER: US/09/183,959
;  CURRENT FILING DATE: 1998-11-02
;  EARLIER APPLICATION NUMBER: 60/064,493
;  EARLIER FILING DATE: No. 6303332ember 5, 1997
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: Microsoft Windows 95
; SEQ ID NO 11
;   LENGTH: 1040
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-183-959-11


  Query Match             52.0%;  Score 13;  DB 4;  Length 1040;
  Best Local Similarity   100.0%;  Pred. No. 92;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     426 ggtgcaggtcagc 438


RESULT  34
US-08-706-216-5/c
; Sequence 5, Application US/08706216
; Patent No. 6140098
;  GENERAL INFORMATION:
;    APPLICANT:  Balasubramanian, Sriram
;    APPLICANT:  Ford, John
;    APPLICANT:  Gorman, Daniel M.
;    APPLICANT:  Zurawski, Gerard
;    TITLE OF INVENTION:  MAMMALIAN PROTEASES; RELATED REAGENTS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DNAX Research Institute
;      STREET:  901 California Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/706,216
;      FILING DATE:  30-AUG-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ching, Edwin P.
;      REGISTRATION NUMBER:  34,090
;      REFERENCE/DOCKET NUMBER:  DX0613
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-852-9196
;      TELEFAX:  415-496-1200
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1173 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1173
US-08-706-216-5


  Query Match             52.0%;  Score 13;  DB 3;  Length 1173;
  Best Local Similarity   100.0%;  Pred. No. 91;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtc 16
           |||||||||||||
Db     418 TCAGGTGCAGGTC 406


RESULT  35
US-08-488-376-17/c
```

```
; Sequence 17, Application US/08488376
; Patent No. 5811524
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-488-376-17


  Query Match             52.0%;  Score 13;  DB 1;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


RESULT  36
US-08-488-376-19/c
; Sequence 19, Application US/08488376
; Patent No. 5811524
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-488-376-19


  Query Match             52.0%;  Score 13;  DB 1;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db     127 AGGTGCAGGTCAG 115


RESULT  37
US-08-634-223-17/c
; Sequence 17, Application US/08634223
; Patent No. 5840298
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/634,223
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-223-17


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


RESULT  38
US-08-634-223-19/c
; Sequence 19, Application US/08634223
; Patent No. 5840298
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,223
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-223-19


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db     127 AGGTGCAGGTCAG 115


RESULT  39
US-08-634-224-17/c
; Sequence 17, Application US/08634224
; Patent No. 5866125
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,224
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
```

```
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-224-17


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


RESULT  40
US-08-634-224-19/c
; Sequence 19, Application US/08634224
; Patent No. 5866125
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,224
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-224-19


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db     127 AGGTGCAGGTCAG 115


RESULT  41
US-08-634-400-17/c
; Sequence 17, Application US/08634400
; Patent No. 5939068
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,400
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-400-17


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


RESULT  42
US-08-634-400-19/c
; Sequence 19, Application US/08634400
; Patent No. 5939068
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,400
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-634-400-19


  Query Match           52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity  100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db     127 AGGTGCAGGTCAG 115


RESULT  43
US-08-635-878-17/c
; Sequence 17, Application US/08635878
; Patent No. 5955364
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/635,878
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/488,376
;      FILING DATE:  07-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Teskin, Robin L.
;      REGISTRATION NUMBER:  35,030
;      REFERENCE/DOCKET NUMBER:  012712-150
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 836-6620
;      TELEFAX:  (703) 836-2021
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1428 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1428
US-08-635-878-17


  Query Match           52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity  100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


RESULT  44
US-08-635-878-19/c
; Sequence 19, Application US/08635878
; Patent No. 5955364
;  GENERAL INFORMATION:
;    APPLICANT:  BRAMS, Peter
;    APPLICANT:  CHAMAT, Soulaima Salim
;    APPLICANT:  PAN, Li-Zhen
;    APPLICANT:  WALSH, Edward E.
;    APPLICANT:  HEARD, Cheryl Janne
;    APPLICANT:  NEWMAN, Roland Anthony
;    TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;    TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;    TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Burns, Doane, Swecker & Mathis
;      STREET:  P.O. Box 1404
;      CITY:  Alexandria
;      STATE:  Virginia
;      COUNTRY:  United States
;      ZIP:  22313-1404
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/635,878
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/488,376
;       FILING DATE:  07-JUN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Teskin, Robin L.
;       REGISTRATION NUMBER:  35,030
;       REFERENCE/DOCKET NUMBER:  012712-150
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6620
;       TELEFAX:  (703) 836-2021
;   INFORMATION FOR SEQ ID NO:  19:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1428 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1428
US-08-635-878-19


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcag 18
           |||||||||||||
Db     127 AGGTGCAGGTCAG 115


RESULT  45
US-08-770-057-17/c
; Sequence 17, Application US/08770057
; Patent No. 5958765
;  GENERAL INFORMATION:
;     APPLICANT:  BRAMS, Peter
;     APPLICANT:  CHAMAT, Soulaima Salim
;     APPLICANT:  PAN, Li-Zhen
;     APPLICANT:  WALSH, Edward E.
;     APPLICANT:  HEARD, Cheryl Janne
;     APPLICANT:  NEWMAN, Roland Anthony
;     TITLE OF INVENTION:  NEUTRALIZING HIGH AFFINITY HUMAN
;     TITLE OF INVENTION:  MONOCLONAL ANTIBODIES SPECIFIC TO RSV F-PROTEIN AND
;     TITLE OF INVENTION:  METHODS FOR THEIR MANUFACTURE AND THERAPEUTIC USE THEREOF
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Burns, Doane, Swecker & Mathis
;       STREET:  P.O. Box 1404
;       CITY:  Alexandria
;       STATE:  Virginia
;       COUNTRY:  United States
;       ZIP:  22313-1404
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/770,057
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/488,376
;       FILING DATE:  07-JUN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Teskin, Robin L.
;       REGISTRATION NUMBER:  35,030
;       REFERENCE/DOCKET NUMBER:  012712-150
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 836-6620
;       TELEFAX:  (703) 836-2021
;   INFORMATION FOR SEQ ID NO:  17:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1428 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1428
US-08-770-057-17


  Query Match             52.0%;  Score 13;  DB 2;  Length 1428;
  Best Local Similarity   100.0%;  Pred. No. 90;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagc 19
           |||||||||||||
Db     126 GGTGCAGGTCAGC 114


Search completed: December 26, 2001, 12:57:44
Job time: 7096 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:10:41 ; Search time 4619.78 Seconds
(without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196F-10
Perfect score: 25
Sequence: 1 tcgtcaggtgcaggtcagcacgttg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 9

Total number of hits satisfying chosen parameters: 587657

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	17	68.0	324	11	BF464374	BF464374 UI-M-CG0p
c	2	17	68.0	413	11	BE862745	BE862745 UI-M-BH0-
c	3	17	68.0	495	10	AI789420	AI789420 uk99a06.y
c	4	17	68.0	641	13	AZ426129	AZ426129 1M0206H03
c	5	17	68.0	767	11	BG703880	BG703880 602686932
c	6	16	64.0	163	10	AA065941	AA065941 ml51a01.r
c	7	16	64.0	288	10	BB516544	BB516544 BB516544
	8	16	64.0	328	11	W54143	W54143 md13c03.r1
	9	16	64.0	338	11	W34110	W34110 mc57c06.r1
c	10	16	64.0	345	10	AV633196	AV633196 AV633196
c	11	16	64.0	368	10	AA110497	AA110497 ml62e03.r
	12	16	64.0	370	10	AW822106	AW822106 uq13c09.x
	13	16	64.0	395	11	BF444068	BF444068 261798 MA
	14	16	64.0	398	13	AQ298728	AQ298728 HS_3110_B
c	15	16	64.0	430	10	AI467556	AI467556 ve36g09.x
	16	16	64.0	544	10	AW892602	AW892602 CM3-NN000
c	17	16	64.0	553	13	AQ530462	AQ530462 RPCI-11-3
c	18	16	64.0	682	13	AZ978539	AZ978539 2M0254K21
c	19	16	64.0	775	10	BE535032	BE535032 601233227
	20	16	64.0	897	13	CNS021CM	AL176719 Tetraodon
c	21	16	64.0	976	11	BG172968	BG172968 602336627
c	22	16	64.0	1011	12	AK010837	AK010837 Mus muscu
c	23	16	64.0	1044	13	CNS078T3	AL434413 T7 end of
c	24	16	64.0	1298	11	BG328084	BG328084 602427134
c	25	15	60.0	180	10	AA699872	AA699872 zj81f04.s
c	26	15	60.0	264	11	H22139	H22139 yl38a03.s1
c	27	15	60.0	279	11	R50061	R50061 yj59c10.s1
c	28	15	60.0	292	10	AI331540	AI331540 fa94b01.y
c	29	15	60.0	293	13	AZ877721	AZ877721 RPCI-23-1
c	30	15	60.0	300	10	BB175346	BB175346 BB175346
c	31	15	60.0	306	11	BG100182	BG100182 ux94g04.y
	32	15	60.0	382	10	AI138808	AI138808 qd98g02.x
c	33	15	60.0	385	10	AW522699	AW522699 UI-R-BO0-
c	34	15	60.0	396	10	AW486040	AW486040 69428 MAR
c	35	15	60.0	407	10	AA062258	AA062258 ml55g08.r
c	36	15	60.0	407	11	BF599674	BF599674 263546 MA
c	37	15	60.0	408	13	AQ085001	AQ085001 HS_2269_A
c	38	15	60.0	409	11	BF653430	BF653430 277257 MA
c	39	15	60.0	416	10	AA146146	AA146146 mq65d12.r
c	40	15	60.0	446	10	AA881231	AA881231 vx10d04.r
c	41	15	60.0	457	11	BF604067	BF604067 269755 MA
c	42	15	60.0	466	11	BI398127	BI398127 UMN45G05
c	43	15	60.0	477	13	AQ921326	AQ921326 RPCI-23-2
c	44	15	60.0	492	11	BE808281	BE808281 213447 MA
c	45	15	60.0	495	13	AZ261270	AZ261270 RPCI-23-4
c	46	15	60.0	496	11	BF652819	BF652819 276504 MA
	47	15	60.0	525	11	BF567735	BF567735 UI-R-BO0-
c	48	15	60.0	529	10	AW290780	AW290780 NXNV046H0
	49	15	60.0	548	13	AZ887515	AZ887515 RPCI-24-1
	50	15	60.0	574	11	BF044929	BF044929 BP250009A
	51	15	60.0	577	11	BG711675	BG711675 pg11n.pk0
	52	15	60.0	580	10	BE262480	BE262480 601145796
	53	15	60.0	589	10	AV593752	AV593752 AV593752
c	54	15	60.0	598	10	AA519851	AA519851 TgESTzz69
c	55	15	60.0	607	13	AZ873604	AZ873604 2M0187J24
	56	15	60.0	620	13	AZ342718	AZ342718 1M0075I20
c	57	15	60.0	624	13	AZ418216	AZ418216 1M0194P10
c	58	15	60.0	639	10	AW342762	AW342762 fi86h10.x
c	59	15	60.0	696	10	AV709400	AV709400 AV709400
c	60	15	60.0	710	13	AZ998889	AZ998889 2M0286J03
c	61	15	60.0	743	10	BE309561	BE309561 601094957
c	62	15	60.0	804	11	BF864835	BF864835 963055C06
c	63	15	60.0	811	11	BG965551	BG965551 602830761
c	64	15	60.0	841	10	BE376456	BE376456 601226248
c	65	15	60.0	863	11	BF966467	BF966467 602287091
	66	15	60.0	946	13	CNS036ON	AL229424 Tetraodon
c	67	15	60.0	984	13	CNS05T79	AL352782 Tetraodon
	68	15	60.0	990	13	CNS024LL	AL180930 Tetraodon
	69	15	60.0	1043	13	CNS05049	AL315090 Tetraodon
c	70	15	60.0	1064	11	BI078629	BI078629 602872730
	71	15	60.0	1069	13	CNS05GGT	AL336278 Tetraodon
c	72	15	60.0	1101	13	CNS0141H	AL103583 Drosophil
c	73	15	60.0	1143	11	BF124974	BF124974 601762117
c	74	14	56.0	104	11	BF855889	BF855889 PM2-FN021
c	75	14	56.0	111	11	BF361607	BF361607 POT-L-101
	76	14	56.0	143	10	BE066289	BE066289 RC3-BT033
	77	14	56.0	162	10	AV010258	AV010258 AV010258
c	78	14	56.0	169	11	BE941904	BE941904 EST421555
	79	14	56.0	185	10	AI314478	AI314478 uj11h11.x
c	80	14	56.0	193	11	BF333453	BF333453 IL3-CI001
	81	14	56.0	197	11	BF412527	BF412527 UI-R-BT1-
c	82	14	56.0	199	10	AI789574	AI789574 ui26d08.r
c	83	14	56.0	200	10	BE218695	BE218695 hv43d11.x
c	84	14	56.0	204	10	AW627429	AW627429 hh82g05.x
c	85	14	56.0	204	11	BF772233	BF772233 CM4-IT004

c	86	14	56.0	210	10	BB254937	BB254937 BB254937
	87	14	56.0	218	11	BI016166	BI016166 PM1-ET025
c	88	14	56.0	226	10	BB427689	BB427689 BB427689
c	89	14	56.0	227	10	BB385773	BB385773 BB385773
c	90	14	56.0	232	11	F17459	F17459 HSPD02001 H
c	91	14	56.0	234	10	BB422516	BB422516 BB422516
c	92	14	56.0	234	10	BB423291	BB423291 BB423291
c	93	14	56.0	234	10	BB426847	BB426847 BB426847
c	94	14	56.0	234	10	BB474068	BB474068 BB474068
c	95	14	56.0	234	10	BB540338	BB540338 BB540338
c	96	14	56.0	235	10	BB527861	BB527861 BB527861
c	97	14	56.0	237	10	BB390976	BB390976 BB390976
c	98	14	56.0	237	10	BB482445	BB482445 BB482445
c	99	14	56.0	239	10	BB362025	BB362025 BB362025
c	100	14	56.0	240	10	BB484401	BB484401 BB484401

ALIGNMENTS

```
RESULT   1
BF464374/c
LOCUS       BF464374      324 bp    mRNA            EST       04-DEC-2000
DEFINITION  UI-M-CG0p-bog-f-09-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
            UI-M-CG0p-bog-f-09-0-UI 3', mRNA sequence.
ACCESSION   BF464374
VERSION     BF464374.1  GI:11533557
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
            clones from RESEARCH GENETICS. It should be noted that Bento Soares
            is generating a small number of additional specialized
            non-redundant arrays of BMAP cDNAs whose availability will be
            considered under appropriate and limited collaborative arrangements
            The following repetitive elements were found in this cDNA sequence:
            102-169, >(GAA)n#Simple_repeat
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-CG0p-bog-f-09-0-UI"
                     /clone_lib="NIH_BMAP_Ret4_S2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_Ret4_S2 library is a subtracted library,
                     ultimately derived from mouse retina tissue libraries at
                     various stages of development. For a detailed description
                     of the library from which this clone was derived, please
                     visit our web site at brainest.eng.uiowa.edu.
                     TAG_SEQ=None found"
BASE COUNT       91 a     84 c     89 g     60 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 11;  Length 324;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           ||||||||||||||||||
Db     324 CGTCAGGTGCAGGTCAG 308


RESULT   2
BE862745/c
LOCUS       BE862745      413 bp    mRNA            EST       29-SEP-2000
DEFINITION  UI-M-BH0-ajf-b-11-0-UI.r1 NIH_BMAP_M_S1 Mus musculus cDNA clone
            UI-M-BH0-ajf-b-11-0-UI 5', mRNA sequence.
ACCESSION   BE862745
VERSION     BE862745.1  GI:10382015
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH0-ajf-b-11-0-UI"
                     /clone_lib="NIH_BMAP_M_S1"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S1 library is a subtracted library derived from
                     a mixture of normalized libraries from ten regions of the
                     mouse brain (cerebellum, brain stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal ganglia, pineal
                     gland, striatum, hipoccampus). The driver used for
                     subtraction consisted of a pool of 20,000 cDNA clones
                     obtained from non-normalized and normalized libraries of
                     these ten regions of the mouse brain."
BASE COUNT       92 a     125 c     112 g     84 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 11;  Length 413;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           ||||||||||||||||||
```

```
Db     17 CGTCAGGTGCAGGTCAG 1


RESULT   3
AI789420/c
LOCUS       AI789420      495 bp    mRNA            EST       02-JUL-1999
DEFINITION  uk99a06.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2064658 5', mRNA sequence.
ACCESSION   AI789420
VERSION     AI789420.1  GI:5337136
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: uk99a06.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:993846
            Seq primer: custom primer used
            High quality sequence stop: 493.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2064658"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      147 a    122 c    141 g     85 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 495;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           |||||||||||||||||
Db     393 CGTCAGGTGCAGGTCAG 377


RESULT   4
AZ426129/c
LOCUS       AZ426129      641 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0206H03R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0206H03 R, DNA sequence.
ACCESSION   AZ426129
VERSION     AZ426129.1  GI:10550142
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0206  row: H  column: 03
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 641.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0206H03"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      156 a    184 c    183 g    118 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 13;  Length 641;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtcag 18
           |||||||||||||||||
Db     212 CGTCAGGTGCAGGTCAG 196


RESULT   5
```

```
BG703880/c
LOCUS       BG703880      767 bp    mRNA            EST       07-MAY-2001
DEFINITION  602686932F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4819446 5',
            mRNA sequence.
ACCESSION   BG703880
VERSION     BG703880.1  GI:13976654
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10722  row: p  column: 07
            High quality sequence stop: 536.
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4819446"
                     /clone_lib="NIH_MGC_95"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.5 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for  full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health).  Note: this is a NIH_MGC Library."
BASE COUNT      139 a    215 c    202 g    211 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 11;  Length 767;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtcag 18
          ||||||||||||||||||
Db    649 CGTCAGGTGCAGGTCAG 633


RESULT   6
AA065941/c
LOCUS       AA065941      163 bp    mRNA            EST       03-FEB-1997
DEFINITION  ml51a01.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
            clone IMAGE:515496 5', mRNA sequence.
ACCESSION   AA065941
VERSION     AA065941.1  GI:1562644
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 163)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:309344
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 144.
FEATURES             Location/Qualifiers
     source          1. .163
                     /organism="Mus musculus"
                     /strain="Inbred CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:515496"
                     /clone_lib="Stratagene mouse testis (#937308)"
                     /sex="males"
                     /tissue_type="testis"
                     /dev_stage="10-12 week old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: testis; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
                     ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       23 a     56 c     49 g     35 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 163;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db     16 GTCAGGTGCAGGTCAG 1


RESULT   7
BB516544/c
LOCUS       BB516544      288 bp    mRNA            EST       28-JUL-2000
DEFINITION  BB516544 RIKEN full-length enriched, 16 days neonate heart Mus
            musculus cDNA clone D830016J11 3' similar to M76601 Mouse alpha
            cardiac myosin heavy chain mRNA, mRNA sequence.
ACCESSION   BB516544
VERSION     BB516544.1  GI:9568002
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 288)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
```

```
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .288
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D830016J11"
                     /clone_lib="RIKEN full-length enriched, 16 days neonate
                     heart"
                     /tissue_type="heart"
                     /dev_stage="16 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       83 a    104 c     57 g     44 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 288;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 cgtcaggtgcaggtca 17
           ||||||||||||||||
Db     188 CGTCAGGTGCAGGTCA 173


RESULT   8
W54143
LOCUS       W54143         328 bp    mRNA            EST       03-JUN-1996
DEFINITION  md13c03.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:368260 5', mRNA sequence.
ACCESSION   W54143
VERSION     W54143.1  GI:1355214
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:229692
            Seq primer: mob.REGA+ET.
FEATURES             Location/Qualifiers
     source          1. .328
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:368260"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT       81 a     66 c     85 g     96 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 328;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtcagc 19
           ||||||||||||||||
Db      86 TCAGGTGCAGGTCAGC 101


RESULT   9
W34110
LOCUS       W34110         338 bp    mRNA            EST       13-MAY-1996
DEFINITION  mc57c06.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA
            clone IMAGE:352618 5', mRNA sequence.
ACCESSION   W34110
VERSION     W34110.1  GI:1316081
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
```

```
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:224418
            Putative full length read
            Seq primer: mob.REGA+ET
            High quality sequence stop: 324.
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:352618"
                     /clone_lib="Soares mouse embryo NbME13.5 14.5"
                     /sex="unknown"
                     /tissue_type="embryo"
                     /dev_stage="13.5-14.5dpc total fetus"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGGAAATTTTTTTTTTTTTTTTTTTTT
                     T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
                     14.5dpc embryos [total RNA provided by  Minoru Ko, Wayne
                     State Univ., from 2 ]; double-stranded cDNA was ligated to
                     Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector. Library went through one round of
                     normalization, and was constructed by Bento Soares and
                     M.Fatima Bonaldo.  "
BASE COUNT       86 a     65 c     84 g    103 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 338;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtcagc 19
           ||||||||||||||||
Db      82 TCAGGTGCAGGTCAGC 97


RESULT  10
AV633196/c
LOCUS       AV633196      345 bp    mRNA            EST       15-DEC-2000
DEFINITION  AV633196 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
            cDNA clone HC017h07_r 5', mRNA sequence.
ACCESSION   AV633196
VERSION     AV633196.1  GI:10776516
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 345)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .345
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="HC017h07_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a medium with bubbling air containing 5% carbon
                     dioxide"
BASE COUNT       62 a    105 c     98 g     79 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 345;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagca 20
           ||||||||||||||||
Db     173 CAGGTGCAGGTCAGCA 158


RESULT  11
AA110497/c
LOCUS       AA110497      368 bp    mRNA            EST       03-FEB-1997
DEFINITION  ml62e03.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
            clone IMAGE:516604 5', mRNA sequence.
ACCESSION   AA110497
VERSION     AA110497.1  GI:1662274
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:310452
            Seq primer: -28m13 rev1 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Mus musculus"
                     /strain="Inbred CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:516604"
                     /clone_lib="Stratagene mouse testis (#937308)"
                     /sex="males"
                     /tissue_type="testis"
                     /dev_stage="10-12 week old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: testis; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
                     ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor
```

```
                sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT       75 a    115 c    106 g     72 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 368;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db     21 GTCAGGTGCAGGTCAG 6


RESULT  12
AW822106
LOCUS       AW822106      370 bp    mRNA            EST       17-MAY-2000
DEFINITION  uq13c09.x1 Ren Stubbs mouse thymus Mus musculus cDNA clone
            IMAGE:2802352 3', mRNA sequence.
ACCESSION   AW822106
VERSION     AW822106.1  GI:7915183
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: uq13c09.y1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1041876
            Seq primer: Primer name ambiguous
            High quality sequence stop: 303.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Mus musculus"
                     /strain="C3H"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2802352"
                     /clone_lib="Ren Stubbs mouse thymus"
                     /sex="mixed"
                     /dev_stage="3 weeks"
                     /lab_host="DH10B"
                     /note="Organ: thymus; Vector: pT7T3D-Pac; Site_1: NotI;
                     Site_2: PacI; 1st strand cDNA was primed with an oligo(dT)
                     primer; double-stranded cDNA was ligated using 5' linker
                     ggccgctat and 3' linker aactggaagcttaatt. Library is
                     size-selected >2.5 kb and average insert size is 3.5 kb.
                     Clones were arrayed from primary plating; non-amplified.
                     Library constructed by X. Ren and L. Stubbs (Lawrence
                     Livermore National Laboratory and DOE Joint Genome
                     Institute, 7000 East Ave, L-453, Livermore, CA 94550)."
BASE COUNT      125 a     57 c     75 g    113 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 370;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db     53 GTCAGGTGCAGGTCAG 68


RESULT  13
BF444068
LOCUS       BF444068      395 bp    mRNA            EST       01-DEC-2000
DEFINITION  261798 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF444068
VERSION     BF444068.1  GI:11504160
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 395)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 96  row: D  column: 1
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .395
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      118 a     66 c    143 g     68 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 395;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagca 20
          ||||||||||||||||
Db    309 CAGGTGCAGGTCAGCA 324


RESULT  14
AQ298728
LOCUS       AQ298728      398 bp    DNA             GSS       15-DEC-1998
DEFINITION  HS_3110_B1_C12_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3110 Col=23 Row=F, DNA sequence.
ACCESSION   AQ298728
VERSION     AQ298728.1  GI:4015907
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 398)
```

```
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3110  row: F  column: 23
            Class: BAC ends
            High quality sequence stop: 398.
FEATURES             Location/Qualifiers
     source          1. .398
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3110 Col=23 Row=F"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       87 a     93 c    102 g    116 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 398;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagcacg 22
          ||||||||||||||||
Db    330 GGTGCAGGTCAGCACG 345


RESULT  15
AI467556/c
LOCUS       AI467556      430 bp    mRNA            EST       09-MAR-1999
DEFINITION  ve36g09.x1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:820288 3', mRNA sequence.
ACCESSION   AI467556
VERSION     AI467556.1  GI:4320893
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:488568
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            High quality sequence stop: 406.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:820288"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      136 a    102 c     85 g    107 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 430;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcagc 19
          ||||||||||||||||
Db    259 TCAGGTGCAGGTCAGC 244


RESULT  16
AW892602
LOCUS       AW892602      544 bp    mRNA            EST       24-MAY-2000
DEFINITION  CM3-NN0004-100300-111-f07 NN0004 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW892602
VERSION     AW892602.1  GI:8056807
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM3-NN0004-100
            300-111-f07&t3=2000-03-10&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 543.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NN0004"
                     /dev_stage="Adult"
                     /note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
```

```
                profiles into  the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."
BASE COUNT      127 a    152 c    160 g    105 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 544;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db    383 GTCAGGTGCAGGTCAG 398


RESULT  17
AQ530462/c
LOCUS       AQ530462      553 bp    DNA             GSS       18-MAY-1999
DEFINITION  RPCI-11-369F11.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-369F11
            , DNA sequence.
ACCESSION   AQ530462
VERSION     AQ530462.1  GI:4842505
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-369F11.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Homo sapiens"
                     /db_xref="GDB:7641442"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-369F11"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      108 a    165 c    157 g    123 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db    204 GTCAGGTGCAGGTCAG 189


RESULT  18
AZ978539/c
LOCUS       AZ978539      682 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0254K21R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0254K21 R, DNA sequence.
ACCESSION   AZ978539
VERSION     AZ978539.1  GI:13849766
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 682)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0254  row: K  column: 21
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 682.
FEATURES             Location/Qualifiers
     source          1. .682
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0254K21"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      173 a    158 c    120 g    231 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 682;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db     55 GTCAGGTGCAGGTCAG 40
```

```
RESULT  19
BE535032/c
LOCUS       BE535032      775 bp    mRNA            EST       09-AUG-2000
DEFINITION  601233227F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3596935 5',
            mRNA sequence.
ACCESSION   BE535032
VERSION     BE535032.1  GI:9763677
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 775)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8775  row: f  column: 08
            High quality sequence stop: 225.
FEATURES             Location/Qualifiers
     source          1. .775
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3596935"
                     /clone_lib="NCI_CGAP_Mam6"
                     /sex="female, virgin"
                     /tissue_type="infiltrating ductal carcinoma"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Jeffrey Green, M.D., NIH"
BASE COUNT      170 a    186 c    221 g    198 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 775;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 cgtcaggtgcaggtca 17
          ||||||||||||||||
Db    201 CGTCAGGTGCAGGTCA 186


RESULT  20
CNS021CM
LOCUS       CNS021CM      897 bp    DNA             GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            225C12 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL176719
VERSION     AL176719.1  GI:7814776
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 897)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 897)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="225C12"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG225BB06LP1~end : T7"
BASE COUNT      118 a    284 c    294 g    182 t     19 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 897;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 gtcaggtgcaggtcag 18
          ||||||||||||||||
Db     16 GTCAGGTGCAGGTCAG 31


RESULT  21
BG172968/c
LOCUS       BG172968      976 bp    mRNA            EST       06-FEB-2001
DEFINITION  602336627F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4459758 5',
            mRNA sequence.
ACCESSION   BG172968
VERSION     BG172968.1  GI:12679671
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 976)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10260  row: e  column: 07
            High quality sequence stop: 514.
FEATURES             Location/Qualifiers
     source          1. .976
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
```

```
                     /clone="IMAGE:4459758"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="3 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT       255 a    250 c    163 g    308 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 976;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db     168 GTCAGGTGCAGGTCAG 153


RESULT  22
AK010837/c
LOCUS       AK010837     1011 bp    mRNA            HTC       05-JUL-2001
DEFINITION  Mus musculus ES cells cDNA, RIKEN full-length enriched library,
            clone:2410198H06, full insert sequence.
ACCESSION   AK010837
VERSION     AK010837.1  GI:12846558
KEYWORDS    CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) ES cells cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2410198H06.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1011)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Methods in enzymology. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (bases 1 to 1011)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome research. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (bases 1 to 1011)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome research. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (bases 1 to 1011)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1011)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,
            Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,
            Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,
            Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,
            Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,
            Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,
            Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .1011
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /db_xref="MGD:MGI:1910427"
                     /db_xref="MGD:MGI:1924031"
                     /clone="2410198H06"
                     /cell_type="ES cells"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
BASE COUNT       255 a    227 c    180 g    349 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 12;  Length 1011;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtcag 18
           ||||||||||||||||
Db     290 GTCAGGTGCAGGTCAG 275


RESULT  23
CNS078T3/c
LOCUS       CNS078T3     1044 bp    DNA             GSS       08-JUL-2001
DEFINITION  T7 end of clone BB0AA023D11 of library BB0AA from strain CBS 4732
            of Pichia angusta, genomic survey sequence.
ACCESSION   AL434413
VERSION     AL434413.1  GI:12217827
KEYWORDS    GSS.
SOURCE      Pichia angusta.
  ORGANISM  Pichia angusta
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1  (bases 1 to 1044)
  AUTHORS   Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
            and Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
            angusta
  JOURNAL   FEBS Lett. 487 (1), 76-81 (2000)
  MEDLINE   20584723
REFERENCE   2  (bases 1 to 1044)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
REFERENCE   3  (bases 1 to 1044)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1044
                     /organism="Pichia angusta"
                     /strain="CBS 4732"
                     /db_xref="taxon:4905"
                     /clone="BB0AA023D11"
                     /clone_lib="BB0AA"
                     /note="end : T7"
     misc_feature    complement(<1. .>999)
                     /note="similar to Saccharomyces cerevisiae ORF YGL026c [
                     TRP5 ; tryptophan synthase ]"
                     /evidence=not_experimental
BASE COUNT      245 a    311 c    228 g    257 t      3 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 1044;
  Best Local Similarity   100.0%;  Pred. No. 1.9e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggtgcaggtcagcacg 22
          ||||||||||||||||
Db    546 GGTGCAGGTCAGCACG 531


RESULT  24
BG328084/c
LOCUS       BG328084     1298 bp    mRNA            EST       27-FEB-2001
DEFINITION  602427134F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4546497 5',
            mRNA sequence.
ACCESSION   BG328084
VERSION     BG328084.1  GI:13134522
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1298)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1233  row: c  column: 10
            High quality sequence stop: 511.
FEATURES             Location/Qualifiers
     source          1. .1298
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4546497"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      316 a    498 c    357 g    127 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 1298;
  Best Local Similarity   100.0%;  Pred. No. 1.8e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tcgtcaggtgcaggtc 16
          ||||||||||||||||
Db    984 TCGTCAGGTGCAGGTC 969


RESULT  25
AA699872/c
LOCUS       AA699872     180 bp    mRNA            EST       19-DEC-1997
DEFINITION  zj81f04.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:461311 3', mRNA sequence.
ACCESSION   AA699872
VERSION     AA699872.1  GI:2702835
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 144.
FEATURES             Location/Qualifiers
     source          1. .180
                     /organism="Homo sapiens"
                     /db_xref="GDB:3752334"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:461311"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
                     liver spleen 1NFLS library.   1st strand cDNA was primed
                     with a Pac I - oligo(dT) primer [5'
                     AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       47 a     57 c     36 g     40 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 10;  Length 180;
  Best Local Similarity   100.0%;  Pred. No. 5.9e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 caggtgcaggtcagc 19
          |||||||||||||||
Db    138 CAGGTGCAGGTCAGC 124


RESULT  26
H22139/c
LOCUS       H22139         264 bp    mRNA            EST       06-JUL-1995
DEFINITION  yl38a03.s1 Soares breast 3NbHBst Homo sapiens cDNA clone
            IMAGE:160492 3', mRNA sequence.
ACCESSION   H22139
VERSION     H22139.1  GI:890834
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 264)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 685
            High quality sequence stops: 241
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 685   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 241.
FEATURES             Location/Qualifiers
     source          1. .264
                     /organism="Homo sapiens"
                     /db_xref="GDB:574535"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:160492"
                     /clone_lib="Soares breast 3NbHBst"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: breast; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of a modified pT7T3 vector (Pharmacia).
                     Library went through one round of normalization to a Cot =
                     20. Library constructed by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT       87 a     47 c     50 g     73 t      7 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 264;
  Best Local Similarity   100.0%;  Pred. No. 5.9e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcag 18
            |||||||||||||||
Db    215 TCAGGTGCAGGTCAG 201


RESULT  27
R50061/c
LOCUS       R50061         279 bp    mRNA            EST       18-MAY-1995
DEFINITION  yj59c10.s1 Soares breast 2NbHBst Homo sapiens cDNA clone
            IMAGE:153042 3', mRNA sequence.
ACCESSION   R50061
VERSION     R50061.1  GI:811963
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 279)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 681
            High quality sequence stops: 257 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 681   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 257.
FEATURES             Location/Qualifiers
     source          1. .279
                     /organism="Homo sapiens"
                     /db_xref="GDB:565317"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:153042"
                     /clone_lib="Soares breast 2NbHBst"
                     /sex="Female"
                     /dev_stage="adult"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: breast; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of a modified pT7T3 vector (Pharmacia).
                     Library went through one round of normalization to a Cot =
                     230. Library constructed by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT       93 a     49 c     54 g     75 t      8 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 279;
  Best Local Similarity   100.0%;  Pred. No. 5.9e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tcaggtgcaggtcag 18
            |||||||||||||||
Db    229 TCAGGTGCAGGTCAG 215


RESULT  28
AI331540/c
LOCUS       AI331540       292 bp    mRNA            EST       28-DEC-1998
DEFINITION  fa94b01.y1 zebrafish fin day1 regeneration Danio rerio cDNA 5',
```

```
            mRNA sequence.
ACCESSION   AI331540
VERSION     AI331540.1  GI:4068099
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Raymond Lee. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 272.
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone_lib="zebrafish fin day1 regeneration"
                     /sex="mixed male and female"
                     /tissue_type="1 day fin regenerates"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; 1st
                     strand cDNA primed with (GA)10ACTAGTCTCGAG(T)18, followed
                     by second strand synthesis, and ligated to 5' adapter (5'
                     )-aattcggcacgag- 3', 3'-gccgtgctc-5'.  cDNA was cloned
                     directionally (EcoRI/XhoI) into Stratagene Zap express
                     lambda phage arms.  Mass invivo excision done to obtain
                     inserts in pBK-CMV phagemid."
BASE COUNT       55 a     83 c     96 g     58 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 ggtgcaggtcagcac 21
           |||||||||||||||
Db     211 GGTGCAGGTCAGCAC 197


RESULT  29
AZ877721/c
LOCUS       AZ877721      293 bp    DNA             GSS       05-MAR-2001
DEFINITION  RPCI-23-195I20.TV RPCI-23 Mus musculus genomic clone RPCI-23-195I20
            , DNA sequence.
ACCESSION   AZ877721
VERSION     AZ877721.1  GI:13195787
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 293)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-195I20.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 195  row: I  column: 20
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .293
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-195I20"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       61 a     74 c     78 g     80 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 293;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtca 17
           |||||||||||||||
Db      55 GTCAGGTGCAGGTCA 41


RESULT  30
BB175346/c
LOCUS       BB175346      300 bp    mRNA            EST       29-JUN-2000
DEFINITION  BB175346 RIKEN full-length enriched, adult male hypothalamus Mus
            musculus cDNA clone A230056E21 3', mRNA sequence.
ACCESSION   BB175346
VERSION     BB175346.1  GI:8834429
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
```

```
          , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
          ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
          , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
          ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE   RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL Unpublished (2000)
COMMENT   Contact: Yoshihide Hayashizaki
          Laboratory for Genome Exploration Research Group, RIKEN Genomic
          Sciences Center(GSC), Yokohama Institute
          The Institute of Physical and Chemical Research (RIKEN)
          1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          Tel: 81-45-503-9222
          Fax: 81-45-503-9216
          Email: genome-res@gsc.riken.go.jp,
          URL:http://genome.gsc.riken.go.jp/
          Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
          ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
          Thermostabilization and thermoactivation of thermolabile enzymes by
          trehalose and its application for the synthesis of full length
          cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
          Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
          Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
          ,Y. and Hayashizaki,Y.
          Automated filtration-based high-throughput plasmid preparation
          system. Genome Res. 9 (5), 463-470 (1999)
          Carninci,P. and Hayashizaki,Y.
          High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
          19-44 (1999)
          Please visit our web site (http://genome.rtc.riken.go.jp) for
          further details.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="A230056E21"
                     /clone_lib="RIKEN full-length enriched, adult male
                     hypothalamus"
                     /sex="male"
                     /tissue_type="hypothalamus"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I. "
BASE COUNT       66 a     81 c     58 g     95 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 300;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tcaggtgcaggtcag 18
           |||||||||||||||
Db      72 TCAGGTGCAGGTCAG 58


RESULT  31
BG100182/c
LOCUS       BG100182      306 bp    mRNA            EST       29-JAN-2001
DEFINITION  ux94g04.y1 McCarrey Eddy type B spermatogonia Mus musculus cDNA
            clone IMAGE:3656262 5', mRNA sequence.
ACCESSION   BG100182
VERSION     BG100182.1  GI:12595499
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1418566
            Seq primer: Primer name ambiguous.
FEATURES             Location/Qualifiers
     source          1. .306
                     /organism="Mus musculus"
                     /strain="CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3656262"
                     /clone_lib="McCarrey Eddy type B spermatogonia"
                     /sex="male"
                     /tissue_type="type B spermatogonia, pooled from multiple
                     mice"
                     /dev_stage="8 day"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pBluescript SK+ (Stratagene
                     ); Site_1: XhoI; Site_2: EcoRI; cDNA oligo dT-primed
                     [5'-(GA)10-ACTAGTCTCGAGTTTTTTTTTTTT-3'] and directionally
                     cloned  using 5' linkers 5'-AATTCGGCACGAG-3' and
                     5'-CTCGTGCCG-3'. Size selection of >400bp material gives
                     average insert size ranging from 1-2 kb. Library was mass
                     excised (from  lambda-UniZAP-XR) and resulting
                     single-stranded phagemids were prepped and tranformed
                     into DH10B. Library contains 96% recombinants.
                     References: J. Androl. 20:635-639 and Gene 25:263-269.
                     Library constructed and donated by J. McCarrey, Ph.D.
                     (Southwest Foundation for Biomedical Research, Dept. of
                     Genetics); excision done by E.M. Eddy, Ph.D. (National
                     Institutes of  Health, National Institute of Environmental
                     Health Sciences).  Original lambda-based library is
                     available through ATCC, catalog #63417."
BASE COUNT       62 a     82 c     93 g     69 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 306;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtca 17
           |||||||||||||||
Db      95 GTCAGGTGCAGGTCA 81


RESULT  32
AI138808
LOCUS       AI138808      382 bp    mRNA            EST       28-OCT-1998
DEFINITION  qd98g02.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1737554
```

```
            3', mRNA sequence.
ACCESSION   AI138808
VERSION     AI138808.1  GI:3644780
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 382)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 593   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 362.
FEATURES             Location/Qualifiers
     source          1. .382
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1737554"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      103 a     84 c     75 g    120 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 10;  Length 382;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 aggtgcaggtcagca 20
           ||||||||||||||||
Db     311 AGGTGCAGGTCAGCA 325
```

```
RESULT  33
AW522699/c
LOCUS       AW522699      385 bp    mRNA            EST       06-MAR-2000
DEFINITION  UI-R-BO0-ahk-c-01-0-UI.s1 UI-R-BO0 Rattus norvegicus cDNA clone
            UI-R-BO0-ahk-c-01-0-UI 3', mRNA sequence.
ACCESSION   AW522699
VERSION     AW522699.1  GI:7165124
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized medulla library cDNA Library Preparation: M.B. Soares
            Lab Clone distribution: clones will be available through Research
            Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BO0-ahk-c-01-0-UI"
                     /clone_lib="UI-R-BO0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The library
                     (UI-R-BO0) is a subtracted library derived from a mixture
                     of the following tissues: thalamus, cerebellum,
                     hypothalamus, medulla, pons, midbrain, cerebral cortex,
                     corpus striatum and hippocampus. For a detailed
                     description of the library from which this clone was
                     derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_LIB=UI-R-BO0
                     TAG_TISSUE=medulla
                     TAG_SEQ=GAACCG"
BASE COUNT       87 a    104 c     98 g     96 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 10;  Length 385;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 gtgcaggtcagcacg 22
           |||||||||||||||
Db     336 GTGCAGGTCAGCACG 322
```

```
RESULT  34
AW486040/c
LOCUS       AW486040      396 bp    mRNA            EST       25-APR-2001
DEFINITION  69428 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW486040
VERSION     AW486040.1  GI:7056146
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
```

```
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 45  row: I  column: 7
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .396
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 4BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 20 and day 40
                     embryos."
BASE COUNT       70 a    134 c    127 g     65 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 396;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db     221 CAGGTGCAGGTCAGC 207


RESULT  35
AA062258/c
LOCUS       AA062258      407 bp    mRNA            EST       03-FEB-1997
DEFINITION  ml55g08.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
            clone IMAGE:515966 5', mRNA sequence.
ACCESSION   AA062258
VERSION     AA062258.1  GI:1556057
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:309814
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 387.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Mus musculus"
                     /strain="Inbred CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:515966"
                     /clone_lib="Stratagene mouse testis (#937308)"
                     /sex="males"
                     /tissue_type="testis"
                     /dev_stage="10-12 week old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: testis; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
                     -5' adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT       88 a    117 c    123 g     79 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 407;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtca 17
           |||||||||||||||
Db      15 GTCAGGTGCAGGTCA 1


RESULT  36
BF599674/c
LOCUS       BF599674      407 bp    mRNA            EST       25-APR-2001
DEFINITION  263546 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF599674
VERSION     BF599674.1  GI:11696393
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 35  row: E  column: 3
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
```

```
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       69 a    139 c    135 g     64 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 407;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db      93 CAGGTGCAGGTCAGC 79


RESULT  37
AQ085001/c
LOCUS       AQ085001      408 bp    DNA             GSS       26-AUG-1998
DEFINITION  HS_2269_A2_B10_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2269 Col=20 Row=C, DNA sequence.
ACCESSION   AQ085001
VERSION     AQ085001.1  GI:3454218
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 408)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2269  row: C  column: 20
            Class: BAC ends
            High quality sequence stop: 408.
FEATURES             Location/Qualifiers
     source          1. .408
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2269 Col=20 Row=C"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      104 a    111 c     78 g    115 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 408;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db     396 CAGGTGCAGGTCAGC 382


RESULT  38
BF653430/c
LOCUS       BF653430      409 bp    mRNA            EST       25-APR-2001
DEFINITION  277257 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF653430
VERSION     BF653430.1  GI:11918562
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 69  row: P  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       80 a    138 c    129 g     62 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db     255 CAGGTGCAGGTCAGC 241


RESULT  39
AA146146/c
LOCUS       AA146146      416 bp    mRNA            EST       18-FEB-1997
DEFINITION  mq65d12.r1 Soares_thymus_2NbMT Mus musculus cDNA clone IMAGE:583607
            5', mRNA sequence.
ACCESSION   AA146146
VERSION     AA146146.1  GI:1715582
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
```

```
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:358255
            Seq primer: -28M13 rev2 from Amersham.
FEATURES             Location/Qualifiers
     source          1. .416
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:583607"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      129 a     83 c    125 g     79 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 416;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 gtgcaggtcagcacg 22
          |||||||||||||||
Db    173 GTGCAGGTCAGCACG 159


RESULT  40
AA881231/c
LOCUS       AA881231      446 bp    mRNA            EST       26-MAR-1998
DEFINITION  vx10d04.r1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:1264039 5', mRNA sequence.
ACCESSION   AA881231
VERSION     AA881231.1  GI:2990541
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:666591
            Seq primer: -28m13 rev2 ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1264039"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      141 a     87 c    131 g     87 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 446;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 gtgcaggtcagcacg 22
          |||||||||||||||
Db    158 GTGCAGGTCAGCACG 144


RESULT  41
BF604067/c
LOCUS       BF604067      457 bp    mRNA            EST       25-APR-2001
DEFINITION  269755 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BF604067
VERSION     BF604067.1  GI:11702307
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
```

```
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 50  row: I  column: 16
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 3BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: Xba1; Site_2: XhoI;
                     Library made from pooled tissue from marrow, alveolar
                     macrophage, ovary, fetal semitendonosus muscle, and fetal
                     longissimus muscle."
BASE COUNT       82 a    152 c    142 g     80 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 457;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db     215 CAGGTGCAGGTCAGC 201


RESULT  42
BI398127/c
LOCUS       BI398127      466 bp    mRNA            EST       09-AUG-2001
DEFINITION  UMN45G05 Canine Brain cDNA Library Canis familiaris cDNA 5', mRNA
            sequence.
ACCESSION   BI398127
VERSION     BI398127.1  GI:15145202
KEYWORDS    EST.
SOURCE      dog.
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Roberts,M.C., Hendrickson,J.A., Hoffmann,D.E., Flickinger,G.H.,
            Rutherford,M.S. and Mickelson,J.R.
  TITLE     University of Minnesota Canine Brain EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Mickelson, J.R.
            Veterinary PathoBiology
            University of Minnesota
            1988 Fitch Avenue, University of Minnesota, St. Paul, MN 55108, USA
            Tel: 612 624 1246
            Fax: 612 625 0204
            Email: micke001@umn.edu
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Canis familiaris"
                     /db_xref="taxon:9615"
                     /clone_lib="Canine Brain cDNA Library"
                     /sex="Male"
                     /note="Organ: Brain; Vector: pSPORT1 (Gibco BRL); Site_1:
                     NotI; Site_2: SalI; Tissue was taken from the frontal,
                     occipital, temporal and parietal  lobes, olfactory bulb,
                     hippocampus, cerebellum, thalamus, hypothalamus,  midbrain
                     , pons, and medulla."
BASE COUNT      112 a     94 c    190 g     66 t      4 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 466;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtca 17
           |||||||||||||||
Db     446 GTCAGGTGCAGGTCA 432


RESULT  43
AQ921326/c
LOCUS       AQ921326      477 bp    DNA             GSS       21-DEC-1999
DEFINITION  RPCI-23-273P15.TV RPCI-23 Mus musculus genomic clone RPCI-23-273P15
            , DNA sequence.
ACCESSION   AQ921326
VERSION     AQ921326.1  GI:6610329
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 273  row: P  column: 15
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-273P15"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT      101 a    107 c    125 g    144 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 477;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtca 17
           |||||||||||||||
Db      68 GTCAGGTGCAGGTCA 54


RESULT  44
BE808281/c
LOCUS       BE808281      492 bp    mRNA            EST       25-APR-2001
DEFINITION  213447 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE808281
VERSION     BE808281.1  GI:10239393
KEYWORDS    EST.
```

```
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
            Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
            ,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
            Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
            Keele,J.W.
  TITLE     Sequence evaluation of four pooled-tissue normalized bovine cDNA
            libraries and construction of a gene index for cattle
  JOURNAL   Genome Res. 11 (4), 626-630 (2001)
  MEDLINE   21180013
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 68  row: P  column: 2
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 2BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, thymus,
                     semitendonosus muscle, longissimus muscle, pancreas,
                     adrenal, and endometrium."
BASE COUNT       96 a    161 c    149 g     85 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 492;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 caggtgcaggtcagc 19
           |||||||||||||||
Db     263 CAGGTGCAGGTCAGC 249


RESULT  45
AZ261270/c
LOCUS       AZ261270      495 bp    DNA             GSS       26-JUL-2000
DEFINITION  RPCI-23-440I8.TV RPCI-23 Mus musculus genomic clone RPCI-23-440I8,
            DNA sequence.
ACCESSION   AZ261270
VERSION     AZ261270.1  GI:9469499
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-440I8.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 440  row: I  column: 8
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-440I8"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).  "
BASE COUNT      106 a    110 c    132 g    147 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 495;
  Best Local Similarity   100.0%;  Pred. No. 5.8e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 gtcaggtgcaggtca 17
           |||||||||||||||
Db      66 GTCAGGTGCAGGTCA 52


Search completed: December 26, 2001, 14:10:45
Job time: 7092 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:50:03 ; Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: US-09-396-196F-2
Perfect score: 25
Sequence: 1 attgtcgcaagtcacagaattattt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 220472

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		25	100.0	1041	6	AR029499	AR029499 Sequence
2		25	100.0	1041	6	AR034916	AR034916 Sequence
3		25	100.0	1084	6	A11530	A11530 BioB gene o
4		25	100.0	1121	6	E00893	E00893 Genomic DNA
5		25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
6		25	100.0	5872	6	A38246	A38246 Sequence 1
7		25	100.0	5872	6	A38251	A38251 Sequence 6
8		25	100.0	5872	6	A93674	A93674 Sequence 1
9		25	100.0	5872	6	A93679	A93679 Sequence 6
10		25	100.0	5872	6	AR101809	AR101809 Sequence
11		25	100.0	5872	6	AR101810	AR101810 Sequence
12		25	100.0	11022	1	AE000180	AE000180 Escherich
13		25	100.0	13501	1	AE005258	AE005258 Escherich
14		25	100.0	297816	1	AP002553	AP002553 Escherich
15		24	96.0	5526	1	AF250776	AF250776 Unculture
16	c	17	68.0	25256	9	AP000954	AP000954 Homo sapi
17	c	17	68.0	32767	2	AC087305	AC087305 Homo sapi
18		17	68.0	60072	2	AC011937	AC011937 Homo sapi
19	c	17	68.0	86757	2	HS54J18	AL133366 Homo sapi
20	c	17	68.0	115551	9	AL353582	AL353582 Human DNA
21	c	17	68.0	147706	8	AC083945	AC083945 Oryza sat
22		17	68.0	150613	2	AP003878	AP003878 Oryza sat
23		17	68.0	155172	2	AC021110	AC021110 Homo sapi
24	c	17	68.0	172704	2	AC092548	AC092548 Oryza sat
25	c	17	68.0	174815	2	AP001390	AP001390 Homo sapi
26	c	17	68.0	178056	2	AC091589	AC091589 Homo sapi
27	c	17	68.0	228516	9	AC002523	AC002523 Homo sapi
28	c	17	68.0	340000	9	AP001690	AP001690 Homo sapi
29	c	16	64.0	605	9	HSAPAF20	AF098889 Homo sapi
30	c	16	64.0	1829	11	G07181	G07181 human STS W
31	c	16	64.0	4602	9	D13644	D13644 Homo sapien
32		16	64.0	5469	1	LLA012388	AJ012388 Lactococc
33	c	16	64.0	43741	9	AL161800	AL161800 Human DNA
34	c	16	64.0	68894	2	AC074073	AC074073 Homo sapi
35	c	16	64.0	76075	9	AL590387	AL590387 Human DNA
36	c	16	64.0	105937	8	AC012679	AC012679 Arabidops
37	c	16	64.0	110000	2	AC044784_1	Continuation (2 of
38		16	64.0	110525	9	AB020878	AB020878 Homo sapi
39		16	64.0	111312	9	AC011248	AC011248 Homo sapi
40	c	16	64.0	119733	9	HSDJ655C5	AL121956 Human DNA
41	c	16	64.0	124103	2	AC008951	AC008951 Homo sapi
42		16	64.0	129010	9	AL159168	AL159168 Human DNA
43	c	16	64.0	139218	2	AC019289	AC019289 Homo sapi
44	c	16	64.0	140002	2	AC092893	AC092893 Homo sapi
45		16	64.0	151206	2	AC093258	AC093258 Homo sapi
46	c	16	64.0	153221	2	AC067998	AC067998 Homo sapi
47	c	16	64.0	155211	2	AC019258	AC019258 Homo sapi
48	c	16	64.0	156184	2	AC041018	AC041018 Homo sapi
49		16	64.0	158408	9	AL391821	AL391821 Human DNA
50		16	64.0	162383	9	AC074045	AC074045 Homo sapi
51		16	64.0	167110	2	AC022621	AC022621 Homo sapi
52		16	64.0	167803	2	AC010816	AC010816 Homo sapi
53		16	64.0	168165	9	AC073913	AC073913 Homo sapi
54	c	16	64.0	169671	2	AC021098	AC021098 Homo sapi
55		16	64.0	170949	2	AL590416	AL590416 Homo sapi
56		16	64.0	175931	2	AC026618	AC026618 Homo sapi
57	c	16	64.0	177142	2	AC026392	AC026392 Homo sapi
58		16	64.0	180251	2	AC010807	AC010807 Homo sapi
59		16	64.0	180776	2	AC032043	AC032043 Homo sapi
60	c	16	64.0	182125	2	AC090105	AC090105 Homo sapi
61		16	64.0	182442	2	AC013283	AC013283 Homo sapi
62		16	64.0	183166	2	AL357554	AL357554 Homo sapi
63	c	16	64.0	183856	2	AC021300	AC021300 Homo sapi
64	c	16	64.0	193970	2	AL357119	AL357119 Homo sapi
65	c	16	64.0	196550	9	AC020734	AC020734 Homo sapi
66		16	64.0	196975	2	AC083848	AC083848 Homo sapi
67	c	16	64.0	204164	9	AC020601	AC020601 Homo sapi
68		16	64.0	208656	2	AC023778	AC023778 Homo sapi
69	c	16	64.0	212134	2	AC024037	AC024037 Homo sapi
70		16	64.0	213510	2	AL359921	AL359921 Homo sapi

c	71	16	64.0	214225	2	AC026112	AC026112 Homo sapi
	72	16	64.0	216378	2	AC090420	AC090420 Homo sapi
c	73	16	64.0	233454	9	AC005703	AC005703 Homo sapi
	74	15	60.0	561	6	AX136747	AX136747 Sequence
c	75	15	60.0	1156	6	AX086780	AX086780 Sequence
c	76	15	60.0	1156	9	HSM801807	AL136839 Homo sapi
	77	15	60.0	1771	6	AR080743	AR080743 Sequence
	78	15	60.0	1771	10	MUSINTESUB	L20788 Mus musculu
	79	15	60.0	1949	10	MMINTEGR01	U34627 Mus musculu
	80	15	60.0	3140	9	Y029487S18	AY029504 Homo sapi
c	81	15	60.0	3210	9	AK024286	AK024286 Homo sapi
c	82	15	60.0	3520	6	AX136409	AX136409 Sequence
c	83	15	60.0	3748	6	AX098224	AX098224 Sequence
c	84	15	60.0	5218	9	AF152313	AF152313 Homo sapi
c	85	15	60.0	5221	9	AF152315	AF152315 Homo sapi
c	86	15	60.0	5233	9	AF152308	AF152308 Homo sapi
c	87	15	60.0	5251	9	AF152312	AF152312 Homo sapi
c	88	15	60.0	5254	9	AF152306	AF152306 Homo sapi
c	89	15	60.0	5254	9	AF152310	AF152310 Homo sapi
c	90	15	60.0	5257	9	AF152307	AF152307 Homo sapi
c	91	15	60.0	5260	9	AF152305	AF152305 Homo sapi
c	92	15	60.0	5260	9	AF152309	AF152309 Homo sapi
c	93	15	60.0	5260	9	AF152311	AF152311 Homo sapi
c	94	15	60.0	5260	9	AF152314	AF152314 Homo sapi
c	95	15	60.0	5260	9	AF152316	AF152316 Homo sapi
c	96	15	60.0	5260	9	AF152317	AF152317 Homo sapi
c	97	15	60.0	5299	9	AF152303	AF152303 Homo sapi
c	98	15	60.0	5431	9	AF152304	AF152304 Homo sapi
	99	15	60.0	5562	9	AB011138	AB011138 Homo sapi
	100	15	60.0	6767	9	AB051358	AB051358 Homo sapi

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     24 ATTGTCGCAAGTCACAGAATTATTT 48


RESULT   2
AR034916
LOCUS       AR034916      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     24 ATTGTCGCAAGTCACAGAATTATTT 48


RESULT   3
A11530
LOCUS       A11530        1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     47 ATTGTCGCAAGTCACAGAATTATTT 71


RESULT   4
E00893
LOCUS       E00893        1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
```

```
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db      65 ATTGTCGCAAGTCACAGAATTATTT 89


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
```

```
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAC
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match           100.0%;  Score 25;  DB 1;  Length 5793;
  Best Local Similarity  100.0%;  Pred. No. 6.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db   2035 ATTGTCGCAAGTCACAGAATTATTT 2059


RESULT   6
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
```

```
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
```

```
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db     140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-05;
```

```
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT  10
AR101809
LOCUS       AR101809     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT  11
AR101810
LOCUS       AR101810     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 6.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db    140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT  12
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data.  Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
```

```
CDS             complement(147. .1430)
                /gene="ybhC"
                /function="orf; Not classified"
                /note="f427; 98 pct identical to fragment YBHC_ECOLI
                SW:P46130 (300 aa) but contains 127 additional C-terminal
                residues"
                /codon_start=1
                /transl_table=11
                /product="putative pectinesterase"
                /protein_id="AAC73859.1"
                /db_xref="GI:1786989"
                /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPACTQGVTHTTIQAAVD
                AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRK
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
```

```
  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db    3516 ATTGTCGCAAGTCACAGAATTATTT 3540


RESULT  13
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
```

```
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
```

```
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLKF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match           100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity  100.0%;  Pred. No. 5.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db    9602 ATTGTCGCAAGTCACAGAATTATTT 9626


RESULT  14
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
```

```
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase alpha subunit"
                     /protein_id="BAB34177.1"
                     /db_xref="GI:13360213"
                     /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                     GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                     LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                     AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                     AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                     LADIGEALKTVLK"
gene                 2781. .3686
                     /gene="ECs0755"
CDS                  2781. .3686
                     /gene="ECs0755"
                     /note="probable transcriptional regulator,similar to
                     transcriptional regulators e.g. glycine cleavage system
                     transcription activator (gcv operon activator) -
                     Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                     identity 31 in 300 aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative transcription regulator"
                     /protein_id="BAB34178.1"
                     /db_xref="GI:13360214"
                     /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                     EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                     PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                     LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                     NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                     EKSDSPHLQQFREWLLAKSVVPQEM"
gene                 complement(3720. .4322)
                     /gene="ECs0756"
CDS                  complement(3720. .4322)
                     /gene="ECs0756"
                     /note="probable cob(I)alamin adenosyltransferase,similar
                     to cob(I)alamin adenosyltransferases (corrinoid
                     adenosyltransferases) e.g. [Escherichia coli]
                     gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative cob(I)alamin adenosyltransferase"
                     /protein_id="BAB34179.1"
                     /db_xref="GI:13360215"
                     /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                     KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                     QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                     IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene                 complement(4332. .5984)
                     /gene="ECs0757"
CDS                  complement(4332. .5984)
                     /gene="ECs0757"
                     /note="probable fumarate hydratase,similar to fumarate
                     hydratases e.g. fumarate hydratase class I, aerobic
                     (fumarase) - Escherichia coli
                     gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                     aa"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative fumarate hydratase"
                     /protein_id="BAB34180.1"
                     /db_xref="GI:13360216"
                     /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                     ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                     NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                     TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                     GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                     EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                     EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                     DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                     QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                     MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene                 complement(6092. .7372)
                     /gene="ECs0758"
CDS                  complement(6092. .7372)
                     /gene="ECs0758"
                     /note="probable transport protein,similar to
                     glutamate/aspartate transport proteins (proton glutamate
                     symport proteins) e.g. [Bacillus stearothermophilus]
                     gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                     aa,also similar to C4-dicarboxylate transpor"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative transport protein"
                     /protein_id="BAB34181.1"
                     /db_xref="GI:13360217"
                     /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                     TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                     NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                     FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                     VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                     GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                     LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                     AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene                 complement(7533. .7850)
                     /gene="ECs0759"
CDS                  complement(7533. .7850)
                     /gene="ECs0759"
                     /note="unknown"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="BAB34182.1"

  Query Match             100.0%;  Score 25;  DB 1;  Length 297816;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db   92334 ATTGTCGCAAGTCACAGAATTATTT 92358


RESULT  15
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
```

```
source          1. .5526
                /organism="uncultured bacterium pCosHE2"
                /db_xref="taxon:143797"
                /clone="pCosHE2"
                /note="unknown organism, cosmid clone derived from
                environmental consortium"
CDS             complement(52. .528)
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /product="hypothetical 17.1 kDa protein in modC-bioA
                intergenic region"
                /protein_id="AAG60577.1"
                /db_xref="GI:12620125"
                /translation="MKLISNDLRDGDKLPRRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
gene            complement(587. .1876)
                /gene="bioA"
CDS             complement(587. .1876)
                /gene="bioA"
                /note="7,8-diaminopelargonic acid
                synthetase-aminotransferase"
                /codon_start=1
                /transl_table=11
                /product="DAPA-aminotransferase BioA"
                /protein_id="AAG60578.1"
                /db_xref="GI:12620126"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
gene            1963. .3003
                /gene="bioB"
CDS             1963. .3003
                /gene="bioB"
                /codon_start=1
                /transl_table=11
                /product="biotin synthase BioB"
                /protein_id="AAG60579.1"
                /db_xref="GI:12620127"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            3000. .4154
                /gene="bioF"
CDS             3000. .4154
                /gene="bioF"
                /note="8-amino-7-oxononanoate synthase"
                /codon_start=1
                /transl_table=11
                /product="KAPA synthetase BioF"
                /protein_id="AAG60580.1"
                /db_xref="GI:12620128"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSKA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGKG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSATQPLIVGDNSRALQL
                AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
gene            4141. .4896
                /gene="bioC"
CDS             4141. .4896
                /gene="bioC"
                /note="reaction step prior to synthesis of pimeloyl-CoA"
                /codon_start=1
                /transl_table=11
                /product="biotin biosynthesis protein BioC"
                /protein_id="AAG60581.1"
                /db_xref="GI:12620129"
                /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
gene            4883. .>5526
                /gene="bioD"
CDS             4883. .>5526
                /gene="bioD"
                /note="DTB synthetase"
                /codon_start=1
                /transl_table=11
                /product="dethiobiotin synthetase BioD"
                /protein_id="AAG60582.1"
                /db_xref="GI:12620130"
                /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT      1274 a   1507 c   1567 g   1178 t
ORIGIN


  Query Match             96.0%;  Score 24;  DB 1;  Length 5526;
  Best Local Similarity   100.0%;  Pred. No. 0.00025;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db    1987 TTGTCGCAAGTCACAGAATTATTT 2010


RESULT  16
AP000954/c
LOCUS       AP000954    25256 bp    DNA             PRI       27-JAN-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B705D10,
            LL56-APP region, complete sequence.
ACCESSION   AP000954
VERSION     AP000954.2  GI:6778729
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:B705D10.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 25256)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 25,256 genomic DNA of 21q21.1-q21.2
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 25256)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
            Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     On Jan 27, 2000 this sequence version replaced gi:6630654.
            Sequence updated (26-Jan-2000).
FEATURES             Location/Qualifiers
     source          1. .25256
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="B705D10"
```

```
                     /map="21q21.1-q21.2"
BASE COUNT     8452 a    4090 c    4292 g    8422 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 9;  Length 25256;
  Best Local Similarity   100.0%;  Pred. No. 3.8;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db    1714 AAGTCACAGAATTATTT 1698


RESULT  17
AC087305/c
LOCUS       AC087305     32767 bp    DNA             HTG       23-DEC-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-238P13 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC087305
VERSION     AC087305.1  GI:11990696
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 32767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-238P13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 32767)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12071
                Center clone name: 238_P_13
            --------------.
            * NOTE: This record contains 40 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      707: contig of 707 bp in length
            *      708 807: gap of      100 bp
            *      808     1527: contig of 720 bp in length
            *     1528 1627: gap of      100 bp
            *     1628     2348: contig of 721 bp in length
            *     2349 2448: gap of      100 bp
            *     2449     3157: contig of 709 bp in length
            *     3158 3257: gap of      100 bp
            *     3258     3973: contig of 716 bp in length
            *     3974 4073: gap of      100 bp
            *     4074     4799: contig of 726 bp in length
            *     4800 4899: gap of      100 bp
            *     4900     5617: contig of 718 bp in length
            *     5618 5717: gap of      100 bp
            *     5718     6450: contig of 733 bp in length
            *     6451 6550: gap of      100 bp
            *     6551     7259: contig of 709 bp in length
            *     7260 7359: gap of      100 bp
            *     7360     8066: contig of 707 bp in length
            *     8067 8166: gap of      100 bp
            *     8167     8907: contig of 741 bp in length
            *     8908 9007: gap of      100 bp
            *     9008     9716: contig of 709 bp in length
            *     9717 9816: gap of      100 bp
            *     9817    10532: contig of 716 bp in length
            *    10533 10632: gap of      100 bp
            *    10633    11354: contig of 722 bp in length
            *    11355 11454: gap of      100 bp
            *    11455    12166: contig of 712 bp in length
            *    12167 12266: gap of      100 bp
            *    12267    12976: contig of 710 bp in length
            *    12977 13076: gap of      100 bp
            *    13077    13789: contig of 713 bp in length
            *    13790 13889: gap of      100 bp
            *    13890    14617: contig of 728 bp in length
            *    14618 14717: gap of      100 bp
            *    14718    15452: contig of 735 bp in length
            *    15453 15552: gap of      100 bp
            *    15553    16276: contig of 724 bp in length
            *    16277 16376: gap of      100 bp
            *    16377    17111: contig of 735 bp in length
            *    17112 17211: gap of      100 bp
            *    17212    17925: contig of 714 bp in length
            *    17926 18025: gap of      100 bp
            *    18026    18746: contig of 721 bp in length
            *    18747 18846: gap of      100 bp
            *    18847    19569: contig of 723 bp in length
            *    19570 19669: gap of      100 bp
            *    19670    20384: contig of 715 bp in length
            *    20385 20484: gap of      100 bp
            *    20485    21190: contig of 706 bp in length
            *    21191 21290: gap of      100 bp
            *    21291    22028: contig of 738 bp in length
            *    22029 22128: gap of      100 bp
            *    22129    22841: contig of 713 bp in length
            *    22842 22941: gap of      100 bp
            *    22942    23653: contig of 712 bp in length
            *    23654 23753: gap of      100 bp
            *    23754    24478: contig of 725 bp in length
            *    24479 24578: gap of      100 bp
            *    24579    25305: contig of 727 bp in length
            *    25306 25405: gap of      100 bp
            *    25406    26143: contig of 738 bp in length
            *    26144 26243: gap of      100 bp
            *    26244    26968: contig of 725 bp in length
            *    26969 27068: gap of      100 bp
            *    27069    27785: contig of 717 bp in length
            *    27786 27885: gap of      100 bp
```

```
            *    27886    28584: contig of 699 bp in length
            *    28585 28684: gap of      100 bp
            *    28685    29467: contig of 783 bp in length
            *    29468 29567: gap of      100 bp
            *    29568    30292: contig of 725 bp in length
            *    30293 30392: gap of      100 bp
            *    30393    31117: contig of 725 bp in length
            *    31118 31217: gap of      100 bp
            *    31218    31944: contig of 727 bp in length
            *    31945 32044: gap of      100 bp
            *    32045    32767: contig of 723 bp in length.
FEATURES             Location/Qualifiers
     source          1. .32767
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-238P13"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT     8045 a   5670 c   5396 g   9330 t   4326 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 32767;
  Best Local Similarity   100.0%;  Pred. No. 3.7;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaattat 23
           |||||||||||||||||
Db   30730 GCAAGTCACAGAATTAT 30714


RESULT  18
AC011937
LOCUS       AC011937    60072 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens clone RP11-16E13, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC011937
VERSION     AC011937.2  GI:7144927
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 60072)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-16E13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 60072)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 3, 2000 this sequence version replaced gi:6056246.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3522
                Center clone name: 16_E_13
            --------------.
            * NOTE: This record contains 68 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      763: contig of 763 bp in length
            *      764 863: gap of      100 bp
            *      864     1659: contig of 796 bp in length
            *     1660 1759: gap of      100 bp
            *     1760     2606: contig of 847 bp in length
            *     2607 2706: gap of      100 bp
            *     2707     3494: contig of 788 bp in length
            *     3495 3594: gap of      100 bp
            *     3595     4391: contig of 797 bp in length
            *     4392 4491: gap of      100 bp
            *     4492     5168: contig of 677 bp in length
            *     5169 5268: gap of      100 bp
            *     5269     6058: contig of 790 bp in length
            *     6059 6158: gap of      100 bp
            *     6159     6968: contig of 810 bp in length
            *     6969 7068: gap of      100 bp
            *     7069     7865: contig of 797 bp in length
            *     7866 7965: gap of      100 bp
            *     7966     8753: contig of 788 bp in length
            *     8754 8853: gap of      100 bp
            *     8854     9663: contig of 810 bp in length
            *     9664 9763: gap of      100 bp
            *     9764    10563: contig of 800 bp in length
            *    10564 10663: gap of      100 bp
            *    10664    11465: contig of 802 bp in length
            *    11466 11565: gap of      100 bp
            *    11566    12365: contig of 800 bp in length
            *    12366 12465: gap of      100 bp
            *    12466    13274: contig of 809 bp in length
            *    13275 13374: gap of      100 bp
            *    13375    14163: contig of 789 bp in length
            *    14164 14263: gap of      100 bp
            *    14264    15064: contig of 801 bp in length
            *    15065 15164: gap of      100 bp
            *    15165    15965: contig of 801 bp in length
            *    15966 16065: gap of      100 bp
            *    16066    16852: contig of 787 bp in length
            *    16853 16952: gap of      100 bp
            *    16953    17775: contig of 823 bp in length
            *    17776 17875: gap of      100 bp
            *    17876    18674: contig of 799 bp in length
            *    18675 18774: gap of      100 bp
            *    18775    19477: contig of 703 bp in length
            *    19478 19577: gap of      100 bp
            *    19578    20375: contig of 798 bp in length
            *    20376 20475: gap of      100 bp
            *    20476    21252: contig of 777 bp in length
            *    21253 21352: gap of      100 bp
            *    21353    22153: contig of 801 bp in length
            *    22154 22253: gap of      100 bp
            *    22254    23063: contig of 810 bp in length
            *    23064 23163: gap of      100 bp
            *    23164    23970: contig of 807 bp in length
            *    23971 24070: gap of      100 bp
            *    24071    24697: contig of 627 bp in length
            *    24698 24797: gap of      100 bp
            *    24798    25395: contig of 598 bp in length
```

```
*    25396 25495: gap of      100 bp
*    25496    26303: contig of 808 bp in length
*    26304 26403: gap of      100 bp
*    26404    27193: contig of 790 bp in length
*    27194 27293: gap of      100 bp
*    27294    27964: contig of 671 bp in length
*    27965 28064: gap of      100 bp
*    28065    28724: contig of 660 bp in length
*    28725 28824: gap of      100 bp
*    28825    29615: contig of 791 bp in length
*    29616 29715: gap of      100 bp
*    29716    30499: contig of 784 bp in length
*    30500 30599: gap of      100 bp
*    30600    31391: contig of 792 bp in length
*    31392 31491: gap of      100 bp
*    31492    32268: contig of 777 bp in length
*    32269 32368: gap of      100 bp
*    32369    33177: contig of 809 bp in length
*    33178 33277: gap of      100 bp
*    33278    34070: contig of 793 bp in length
*    34071 34170: gap of      100 bp
*    34171    34978: contig of 808 bp in length
*    34979 35078: gap of      100 bp
*    35079    35870: contig of 792 bp in length
*    35871 35970: gap of      100 bp
*    35971    36751: contig of 781 bp in length
*    36752 36851: gap of      100 bp
*    36852    37644: contig of 793 bp in length
*    37645 37744: gap of      100 bp
*    37745    38573: contig of 829 bp in length
*    38574 38673: gap of      100 bp
*    38674    39462: contig of 789 bp in length
*    39463 39562: gap of      100 bp
*    39563    40367: contig of 805 bp in length
*    40368 40467: gap of      100 bp
*    40468    41259: contig of 792 bp in length
*    41260 41359: gap of      100 bp
*    41360    42171: contig of 812 bp in length
*    42172 42271: gap of      100 bp
*    42272    43084: contig of 813 bp in length
*    43085 43184: gap of      100 bp
*    43185    43981: contig of 797 bp in length
*    43982 44081: gap of      100 bp
*    44082    44891: contig of 810 bp in length
*    44892 44991: gap of      100 bp
*    44992    45788: contig of 797 bp in length
*    45789 45888: gap of      100 bp
*    45889    46671: contig of 783 bp in length
*    46672 46771: gap of      100 bp
*    46772    47570: contig of 799 bp in length
*    47571 47670: gap of      100 bp
*    47671    48460: contig of 790 bp in length
*    48461 48560: gap of      100 bp
*    48561    49352: contig of 792 bp in length
*    49353 49452: gap of      100 bp
*    49453    50256: contig of 804 bp in length
*    50257 50356: gap of      100 bp
*    50357    51147: contig of 791 bp in length
*    51148 51247: gap of      100 bp
*    51248    52076: contig of 829 bp in length
*    52077 52176: gap of      100 bp
*    52177    52976: contig of 800 bp in length
*    52977 53076: gap of      100 bp
*    53077    53891: contig of 815 bp in length
*    53892 53991: gap of      100 bp
*    53992    54778: contig of 787 bp in length
*    54779 54878: gap of      100 bp
*    54879    55683: contig of 805 bp in length
*    55684 55783: gap of      100 bp
*    55784    56484: contig of 701 bp in length
*    56485 56584: gap of      100 bp
*    56585    57369: contig of 785 bp in length
*    57370 57469: gap of      100 bp
*    57470    58276: contig of 807 bp in length
*    58277 58376: gap of      100 bp
*    58377    59169: contig of 793 bp in length
*    59170 59269: gap of      100 bp
*    59270    60072: contig of 803 bp in length.
FEATURES             Location/Qualifiers
     source          1. .60072
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-16E13"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    15923 a   9785 c  10180 g  17174 t   7010 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 60072;
  Best Local Similarity   100.0%;  Pred. No. 3.5;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db    9213 AAGTCACAGAATTATTT 9229


RESULT  19
HS54J18/c
LOCUS       HS54J18     86757 bp    DNA             HTG       28-SEP-2000
DEFINITION  Homo sapiens chromosome 21 clone RPCIP704J1854 map 21q21, ***
            SEQUENCING IN PROGRESS ***, 18 unordered pieces.
ACCESSION   AL133366
VERSION     AL133366.2  GI:10432568
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 86757)
  AUTHORS   Nordsiek,G., Conrad,A., Dose,S., Grimm,M., Hildmann,T.,
            Hornischer,K., Loehnert,T.H., Scharfe,M., Schoen,O., Yaspo,M.L. and
            Bloecker,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) GBF, Dept. of Genome Analysis, Mascheroder
            Weg 1, D-38124 Braunschweig, Germany, E-mail: hornischer@gbf.de,
            nordsiek@gbf.de, bloecker@gbf.de
COMMENT     On Sep 29, 2000 this sequence version replaced gi:6981908.
            Collaborators:
            GBF, Dept. of Genome Analysis
            Mascheroder Weg 1, D-38124 Braunschweig, Germany
            and
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestrasse 73, D-14195 Berlin-Dahlem, Germany
            All annotations in this database entry are developed by
            computational tools. It is therefore not explicitly noted in the
            feature lines that evidence is not experimental.
            PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
            ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
            +   Analysis and annotation were performed with the automatic    +
            +          'first-pass' annotation and submission tool           +
            +             'AnnoMitter' (Hornischer & Bloecker).              +
            +                 Programs used by 'AnnoMitter':                 +
            ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
            > 'Tandem Repeats': GDE 2.2 option 'tandem'
            .   Minimum length 2 bp; Maximum length 20 bp; Score threshold 20 .
              Treat N's as mismatches? YES; Allow uniform consensi? NO >
            'Inverted Repeats': GDE 2.2 option 'inverted'
            > 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
            Islands': GDE 2.2 option 'cpg'
            .   CpG island region size 100 bp;
            .   Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
            e-PCR (Schuler)
            .   Margin: 50; Number of mismatches allowed: 0; Word size: 7 .
            STS database: 'dbSTS markers'.
            * NOTE: This is a 'working draft' sequence. It currently
```

```
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    13423: contig of 13423 bp in length
            *    13424 13523: gap of      100 bp
            *    13524    15096: contig of 1573 bp in length
            *    15097 15196: gap of      100 bp
            *    15197    16448: contig of 1252 bp in length
            *    16449 16548: gap of      100 bp
            *    16549    19018: contig of 2470 bp in length
            *    19019 19118: gap of      100 bp
            *    19119    24811: contig of 5693 bp in length
            *    24812 24911: gap of      100 bp
            *    24912    34477: contig of 9566 bp in length
            *    34478 34577: gap of      100 bp
            *    34578    37261: contig of 2684 bp in length
            *    37262 37361: gap of      100 bp
            *    37362    41884: contig of 4523 bp in length
            *    41885 41984: gap of      100 bp
            *    41985    45057: contig of 3073 bp in length
            *    45058 45157: gap of      100 bp
            *    45158    46158: contig of 1001 bp in length
            *    46159 46258: gap of      100 bp
            *    46259    50846: contig of 4588 bp in length
            *    50847 50946: gap of      100 bp
            *    50947    53362: contig of 2416 bp in length
            *    53363 53462: gap of      100 bp
            *    53463    55903: contig of 2441 bp in length
            *    55904 56003: gap of      100 bp
            *    56004    64319: contig of 8316 bp in length
            *    64320 64419: gap of      100 bp
            *    64420    67502: contig of 3083 bp in length
            *    67503 67602: gap of      100 bp
            *    67603    71620: contig of 4018 bp in length
            *    71621 71720: gap of      100 bp
            *    71721    74980: contig of 3260 bp in length
            *    74981 75080: gap of      100 bp
            *    75081    86757: contig of 11677 bp in length.
FEATURES             Location/Qualifiers
     source          1. .86757
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21"
                     /clone="RPCIP704J1854"
     repeat_region   4088. .4150
                     /note="homology = 73.0%, counts = 7"
                     /rpt_family="tttttatta repeat"
                     /rpt_type=TANDEM
     repeat_region   6744. .6807
                     /note="homology = 78.1%, counts = 8"
                     /rpt_family="tttttatt repeat"
                     /rpt_type=TANDEM
     satellite       6766. .6792
                     /note="TTTA repeat"
     satellite       6793. .6804
                     /note="TAT repeat"
     satellite       7177. .7192
                     /note="AACAA repeat"
     satellite       7426. .7440
                     /note="AATTA repeat"
     repeat_region   8984. .9033
                     /note="IR1, 92% complementary to IR1' (10757. .10806)"
                     /rpt_type=INVERTED
     satellite       9075. .9094
                     /note="AAAC repeat"
     repeat_region   10757. .10806
                     /note="IR1', 92% complementary to IR1 (8984. .9033)"
                     /rpt_type=INVERTED
     satellite       12276. .12290
                     /note="TTG repeat"
     satellite       13116. .13129
                     /note="ATAA repeat"
     repeat_region   15197. .15238
                     /note="IR2, 97% complementary to IR2' (15480. .15521)"
                     /rpt_type=INVERTED
     repeat_region   15480. .15521
                     /note="IR2', 97% complementary to IR2 (15197. .15238)"
                     /rpt_type=INVERTED
     satellite       22885. .22929
                     /note="TG repeat"
     repeat_region   22885. .22928
                     /note="homology = 100.0%, counts = 22"
                     /rpt_family="tg repeat"
                     /rpt_type=TANDEM
     repeat_region   25803. .25870
                     /note="homology = 88.2%, counts = 34"
                     /rpt_family="at repeat"
                     /rpt_type=TANDEM
     repeat_region   25819. .25843
                     /note="IR3, 92% complementary to IR3' (25846. .25870)"
                     /rpt_type=INVERTED
     satellite       25828. .25870
                     /note="TA repeat"
     repeat_region   25846. .25870
                     /note="IR3', 92% complementary to IR3 (25819. .25843)"
                     /rpt_type=INVERTED
     satellite       26732. .26745
                     /note="CAA repeat"
     satellite       35152. .35163
                     /note="TC repeat"
     satellite       38666. .38678
                     /note="TG repeat"
     satellite       40287. .40299
                     /note="ATTT repeat"
     repeat_region   42411. .42436
                     /note="IR4, 84% complementary to IR4' (44122. .44147)"
                     /rpt_type=INVERTED
     repeat_region   44116. .44245
                     /note="homology = 63.1%, counts = 10"
                     /rpt_family="aaaaaaaaaaaat repeat"
                     /rpt_type=TANDEM
     repeat_region   44122. .44147
                     /note="IR4', 84% complementary to IR4 (42411. .42436)"
                     /rpt_type=INVERTED
     satellite       44370. .44383
                     /note="AAAAT repeat"
     repeat_region   46995. .47082
                     /note="homology = 68.2%, counts = 8"
                     /rpt_family="gaaaaaaaaga repeat"
                     /rpt_type=TANDEM
     repeat_region   47093. .47127
                     /note="IR5, 82% complementary to IR5' (47138. .47173)"
                     /rpt_type=INVERTED
     repeat_region   47114. .47175
                     /note="homology = 85.5%, counts = 31"
                     /rpt_family="ta repeat"
                     /rpt_type=TANDEM
     repeat_region   47138. .47173
                     /note="IR5', 82% complementary to IR5 (47093. .47127)"
                     /rpt_type=INVERTED
     satellite       47159. .47175
                     /note="AT repeat"
     repeat_region   47864. .47953
                     /note="homology = 66.7%, counts = 10"
                     /rpt_family="agaaaaaaa repeat"
                     /rpt_type=TANDEM
     repeat_region   48093. .48128
                     /note="homology = 83.3%, counts = 18"
                     /rpt_family="tc repeat"
                     /rpt_type=TANDEM
     repeat_region   48328. .48366
```

```
                     /note="IR6, 82% complementary to IR6' (49401. .49440)"
                     /rpt_type=INVERTED
     repeat_region   49049. .49084
                     /note="homology = 88.9%, counts = 9"
                     /rpt_family="aaat repeat"
                     /rpt_type=TANDEM
     satellite       49049. .49069

  Query Match             68.0%;  Score 17;  DB 2;  Length 86757;
  Best Local Similarity   100.0%;  Pred. No. 3.4;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db   65420 AAGTCACAGAATTATTT 65404


RESULT  20
AL353582/c
LOCUS       AL353582   115551 bp    DNA             PRI       22-MAR-2001
DEFINITION  Human DNA sequence from clone CTD-2190D5 on chromosome 20 Contains
            ESTs, STSs and GSSs, complete sequence.
ACCESSION   AL353582
VERSION     AL353582.16  GI:10443412
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 115551)
  AUTHORS   Skuce,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 1, 2000 this sequence version replaced gi:9988285.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            CTD-2190D5 is from the CalTech genomic sperm BAC library D. VECTOR:
            pBeloBAC11
            IMPORTANT: This sequence is not the entire insert of clone
            CTD-2190D5 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP4-568C11 is at 115452 in this
            sequence. The true right end of clone RP5-966J20 is at 100 in this
            sequence. This sequence was finished as follows unless otherwise
            noted: all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
FEATURES             Location/Qualifiers
     source          1. .115551
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="CTD-2190D5"
                     /clone_lib="CIT-HSP-D1"
     repeat_region   1. .260
                     /note="10 copies 26 mer 88% conserved"
     repeat_region   1400. .1691
                     /note="AluSg repeat: matches 1. .292 of consensus"
     repeat_region   1692. .1818
                     /note="L1PA16 repeat: matches 6020. .6157 of consensus"
     repeat_region   1878. .2076
                     /note="L1MC4 repeat: matches 7699. .7977 of consensus"
     repeat_region   2719. .3328
                     /note="305 copies 2 mer cc 52% conserved"
     repeat_region   2720. .2777
                     /note="2 copies 29 mer 94% conserved"
     repeat_region   2905. .3078
                     /note="2 copies 87 mer 92% conserved"
     repeat_region   2955. .3494
                     /note="6 copies 90 mer 85% conserved"
     repeat_region   2985. .3484
                     /note="50 copies 10 mer ccccctcctg 60% conserved"
     repeat_region   3001. .3510
                     /note="17 copies 30 mer 89% conserved"
     repeat_region   3496. .3625
                     /note="13 copies 10 mer ccccttcctg 63% conserved"
     repeat_region   3512. .3631
                     /note="4 copies 30 mer 87% conserved"
     repeat_region   4283. .4377
                     /note="MER5A repeat: matches 63. .167 of consensus"
     repeat_region   4459. .4588
                     /note="65 copies 2 mer tg 77% conserved"
     repeat_region   4460. .4579
                     /note="4 copies 30 mer 80% conserved"
     repeat_region   4591. .4800
                     /note="7 copies 30 mer 77% conserved"
     repeat_region   4592. .4861
                     /note="3 copies 90 mer 71% conserved"
     repeat_region   4619. .4800
                     /note="7 copies 26 mer 79% conserved"
     repeat_region   8042. .8542
                     /note="LTR45 repeat: matches 1. .513 of consensus"
     repeat_region   8945. .9067
                     /note="MER5B repeat: matches 53. .177 of consensus"
     repeat_region   10365. .10706
                     /note="L2 repeat: matches 2350. .2674 of consensus"
     repeat_region   10990. .11142
                     /note="MLT1G repeat: matches 121. .283 of consensus"
     repeat_region   11228. .11322
                     /note="MLT1E repeat: matches 468. .566 of consensus"
     repeat_region   12021. .12363
                     /note="MER46B repeat: matches 1. .236 of consensus"
     repeat_region   13655. .13678
                     /note="12 copies 2 mer ag 95% conserved"
     repeat_region   13942. .14438
                     /note="L1MB5 repeat: matches 5679. .6174 of consensus"
     repeat_region   15351. .15666
                     /note="2 copies 158 mer 88% conserved"
     misc_feature    complement(15754. .16330)
                     /note="match: GSS: Em:AQ394110"
     repeat_region   15772. .16057
                     /note="MLT1E repeat: matches 267. .567 of consensus"
     repeat_region   16283. .16653
                     /note="MLT1A1 repeat: matches 1. .352 of consensus"
     repeat_region   16889. .17056
                     /note="MIR repeat: matches 59. .229 of consensus"
     repeat_region   17607. .17725
                     /note="AluSg/x repeat: matches 186. .302 of consensus"
     repeat_region   18349. .18558
                     /note="21 copies 10 mer atacacacat 68% conserved"
     repeat_region   18353. .18552
                     /note="100 copies 2 mer at 62% conserved"
     repeat_region   18357. .18564
                     /note="8 copies 26 mer 69% conserved"
     repeat_region   18358. .18567
                     /note="7 copies 30 mer 68% conserved"
     misc_feature    complement(18438. .18827)
                     /note="match: GSS: Em:AQ581588"
```

```
misc_feature    complement(18532. .18952)
                /note="match: GSS: Em:AQ633379"
repeat_region   18813. .18944
                /note="FLAM_A repeat: matches 1. .133 of consensus"
misc_feature    complement(19334. .19731)
                /note="match: GSS: Em:AQ122529"
repeat_region   19472. .19617
                /note="MIR repeat: matches 78. .218 of consensus"
repeat_region   20285. .20513
                /note="MLT1D repeat: matches 320. .501 of consensus"
repeat_region   20695. .20832
                /note="MIR repeat: matches 53. .189 of consensus"
repeat_region   20924. .21387
                /note="MLT1D repeat: matches 2. .505 of consensus"
repeat_region   21818. .21867
                /note="MER91A repeat: matches 1. .48 of consensus"
repeat_region   22143. .22304
                /note="MIR repeat: matches 31. .192 of consensus"
repeat_region   22339. .22584
                /note="MLT1A2 repeat: matches 1. .225 of consensus"
repeat_region   22585. .23315
                /note="LTR39 repeat: matches 1. .794 of consensus"
repeat_region   23316. .23455
                /note="MLT1A2 repeat: matches 225. .370 of consensus"
repeat_region   24413. .24566
                /note="MIR repeat: matches 29. .200 of consensus"
misc_feature    24745. .25217
                /note="match: GSS: Em:AQ403530"
misc_feature    24758. .25162
                /note="match: GSS: Em:AQ435639"
repeat_region   25811. .25956
                /note="MIR repeat: matches 54. .232 of consensus"
repeat_region   27482. .27631
                /note="5 copies 30 mer 77% conserved"
repeat_region   27497. .27631
                /note="9 copies 15 mer 79% conserved"
misc_feature    27635
                /note="Tandem repeat. Forced join. Gap size estimated to
                be approximately 320bp by HindIII, SacI and SspI
                restriction enzyme digest data."
repeat_region   28011. .28380
                /note="MSTB repeat: matches 1. .426 of consensus"
repeat_region   29147. .29602
                /note="LTR24 repeat: matches 1. .490 of consensus"
repeat_region   30051. .30224
                /note="2 copies 87 mer 92% conserved"
repeat_region   32247. .32977
                /note="L1PA16 repeat: matches 5423. .6154 of consensus"
repeat_region   33453. .33754
                /note="LTR8 repeat: matches 390. .691 of consensus"
repeat_region   34281. .34449
                /note="BUR1 repeat: matches 8180. .8350 of consensus"
repeat_region   34808. .34911
                /note="4 copies 26 mer 78% conserved"
repeat_region   34811. .34902
                /note="46 copies 2 mer tg 78% conserved"
repeat_region   34813. .34902
                /note="3 copies 30 mer 84% conserved"
repeat_region   34821. .34910
                /note="9 copies 10 mer tgtgtatgta 77% conserved"
repeat_region   34914. .34991
                /note="3 copies 26 mer 79% conserved"
repeat_region   34916. .34975
                /note="6 copies 10 mer tgtgtgtgta 78% conserved"
repeat_region   35201. .35320
                /note="MER94 repeat: matches 1. .123 of consensus"
repeat_region   35497. .35835
                /note="L1PA8 repeat: matches 5819. .6162 of consensus"
repeat_region   35836. .36095
                /note="AluY repeat: matches 39. .304 of consensus"
repeat_region   36096. .40074
                /note="L1PA8 repeat: matches 1897. .5819 of consensus"
repeat_region   40075. .40337
```

```
  Query Match           68.0%;  Score 17;  DB 9;  Length 115551;
  Best Local Similarity  100.0%;  Pred. No. 3.3;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db   72877 AAGTCACAGAATTATTT 72861


RESULT  21
AC083945/c
LOCUS       AC083945   147706 bp    DNA             PLN       15-JUN-2001
DEFINITION  Oryza sativa strain Nipponbare chromosome X clone OSJNBa0058E19,
            complete sequence.
ACCESSION   AC083945
VERSION     AC083945.3  GI:12061428
KEYWORDS    HTG.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 147706)
  AUTHORS   Spiegel,L.A., King,L., Kirchoff,K.A., de la Bastide,M.,
            Preston,R.R., Nascimento,L.U., Vil,M.D., Baker,J.P., Miller,B.,
            Cunnius,D.M., Kuit,K.H., Rodriguez,S., Santos,L., Zutavern,T.,
            Balija,V.S., Shah,R.S., Bahret,A., Bal,H.P., O'Shaughnessy,A.,
            Dedhia,N.N. and McCombie,W.R.
  TITLE     Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
            Clone OSJNBa0058E19, Complete Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 147706)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2000) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
REFERENCE   3  (bases 1 to 147706)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2001) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
            NY 11724, USA
REFERENCE   4  (bases 1 to 147706)
  AUTHORS   Bal,H.P., Spiegel,L.A., King,L., Kirchoff,K.A., de la Bastide,M.,
            Preston,R.R., Nascimento,L.U., Vil,M.D., Baker,J.P., Miller,B.,
            Cunnius,D.M., Kuit,K.H., Rodriguez,S., Santos,L., Zutavern,T.,
            Balija,V.S., Shah,R.S., Bahret,A., O'Shaughnessy,A., Dedhia,N.N.
            and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2001) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratory, 1, Bungtown Road, Cold Spring Harbor, NY
            11743, USA
  REMARK    Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
            Clone OSJNBa0058E19, Complete Sequence.
REFERENCE   5  (bases 1 to 147706)
  AUTHORS   Bal,H.P., Spiegel,L.A., King,L., Kirchoff,K.A., de la Bastide,M.,
            Preston,R.R., Nascimento,L.U., Vil,M.D., Baker,J.P., Miller,B.,
            Cunnius,D.M., Kuit,K.H., Rodriguez,S., Santos,L., Zutavern,T.,
            Balija,V.S., Shah,R.S., Bahret,A., O'Shaughnessy,A., Dedhia,N.N.
            and McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2001) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratory, 1, Bungtown Road, Cold Spring Harbor, NY
            11743, USA
  REMARK    Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
            Clone OSJNBa0058E19, Complete Sequence.
COMMENT     On Jan 10, 2001 this sequence version replaced gi:11038452.
            This sequence was finished as follows unless otherwise noted:  all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
```

```
        as compressions and repeats; all regions were covered by at least
        one plasmid subclone or more than one M13 subclone; and the
        assembly was confirmed by restriction digest.
FEATURES             Location/Qualifiers
     source          1. .147706
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
                     /chromosome="X"
                     /clone="OSJNBa0058E19"
                     /clone_lib="HindIII"
     gene            58. .5309
                     /gene="OSJNBa0058E19.1"
                     /note="Histidine kinase-like protein"
     CDS             join(58. .156,495. .1156,1654. .2157,2536. .2693,2853. .2957,
                     4891. .5136,5200. .5309)
                     /gene="OSJNBa0058E19.1"
                     /codon_start=1
                     /product="Histidine-kinase like protein"
                     /protein_id="AAK13126.1"
                     /db_xref="GI:13129468"
                     /translation="MAVLVSDQVPDVLIGDPWRFRQIITNLVGNSMKFTEQGHIFIRV
                     HLIEEVKRKMEALDDTSPENIEVTANSKNTMPYNTLSGLEVANNRKTLESFRMFKDSS
                     DAIDSVNLLVTVEDTGIGITKDAQTRIFTPFMQADGSTSRTYGGTGIGLSITKRLVEL
                     MGGEIGFVSKPGVSSTFSFTAIFKENRKDPGDIKRYCPEPTPPDFQGMRALVVDGRCA
                     RAEVTMYHLRRLGIQCDLAATSESALSALLESCNSSVKSSLNMVLVDKEAWGEDSGLA
                     FFRCLIDLRLKGTLKSWQTMPKFFLLAGSITPADSDCLRLAGYSNSIRKPLRLSTVAA
                     CLSKALGVGLTGRRSRDNSLVLRSVLTGKNILVVDDNAVNRIVAAGALKKYGAIVTCV
                     DSGKEAISRLQPPHKFDACFMDVQMPEMDGFEATRLVRSVESKINDTIQAGEVSSEIY
                     GNKAHWHVPILAMTADVIQATFEGSSAKNPFGDNNIAKQKSYAVNREQKFDWTAIYVS
                     FEPTHPDSNTFNTGRAGPDRGRAVQPHWAPQIVGAPPHPPRSLYAAVAQVTGANMLCN
                     DTEQKLQTQRRSQFLPMRRDCRRTALHDSSRVVVVVGGGGGTSVRGAVTPWSSAIDCN
                     RIC"
     repeat_region   4968. .7539
                     /note="Micropon4 Oryza sativa gene, repeat sequence
                     Micropon"
     gene            complement(7197. .10518)
                     /gene="OSJNBa0058E19.2"
                     /note="Putative protein kinase Xa21"
     CDS             complement(join(7197. .7319,7361. .7668,7886. .8184,
                     8255. .8972,9073. .9099,9272. .9425,10121. .10354,
                     10369. .10518))
                     /gene="OSJNBa0058E19.2"
                     /codon_start=1
                     /product="Putative protein kinase Xa21"
                     /protein_id="AAK13127.1"
                     /db_xref="GI:13129469"
                     /translation="MNKLSLALASYLILFLAAAATAATTRCTFEIVVKTDGRRNAGTD
                     ARVSLQGPDADGRQPGVVGPDGRRPRLLREGEPRPVQGRRRLHAVGALQHGAHLRRLW
                     EQARVVRELRDGDAARAGETPVDDAPVANGPKNFAGLWRRSHHQPTTTYWVGKSTPKV
                     PQLVIMIQKLPRTAKPDTAGKPTGGEEVDLSSGAAAAASAAGRRCTSPTSHRARARSG
                     CAPSSPGSGSSRAAVQLRRVSRGCALFVYRAAEGASSTPTPPEEQTTRAAAAAVASPL
                     PALNLTNRGLIGQMSPLGNLTFLKFLFVPANSFTGEIPQSFGNMHHLQIIYLSNNTLQ
                     RRMPNLANCSNLKVLWLNGNNLVGQIPADLPQRFQSLQLSVNSLTGPIPVSVAKITTL
                     KRFSCLYNNIDGSIPNDFAKLPGLVYLHLGAKIIGSGELPSNIGNSLPDLQKFQLGGN
                     FFYGHIPNSLTNASKLNLIDIVSINSFTGVVPRSIGKLTKLSWLNLELNKFHAHSQKD
                     LEFMNSLANCTELQMSSIYGGVGGELAPSAARYTKSAHPRENPVSASKPNGPPSSSPN
                     PAKKARSRSALAPGATLVTYTRLPAAAAAPEDGSAPAVRGELDGDGAAGDDDAAAAAT
                     HLAERTPAQRSREVREKGEKRGNRKVSTWHPDMWGPHGSHADSAAM"
     repeat_region   7923. .9998
                     /note="Retrosat2 Ty3-Gypsy_retroelement"
     gene            18596. .19460
                     /gene="OSJNBa0058E19.3"
                     /note="Hypothetical protein"
     CDS             join(18596. .18726,18794. .18896,18917. .18991,19260. .19460)
                     /gene="OSJNBa0058E19.3"
                     /codon_start=1
                     /product="Hypothetical protein"
                     /protein_id="AAK13128.1"
                     /db_xref="GI:13129470"
                     /translation="MNLIVVQIRKMKSLFLLHSISSKAAMGLWPSARRCRRQMVTPLG
                     GHRSSASGESEQRMFSGGCACRAIDWMYPKGCMHGTHRSSDEVRVGLDSDDDAEDVPS
                     ALYLLHSNRNRRRDLVAAVHCVRSGAPAGEVAFPPNHCMIEAEIRGDGTGIERRRWNT
                     REKETIAAQ"
     gene            complement(22790. .40376)
                     /gene="OSJNBa0058E19.4"
                     /note="Putative gag-pol polyprotein"
     CDS             complement(join(22790. .23208,24472. .24598,24918. .25000,
                     25046. .25158,26244. .26431,26874. .26955,27044. .27126,
                     28396. .28467,28767. .28900,29535. .29647,30186. .30251,
                     30344. .30686,31565. .31711,32362. .32410,33038. .33286,
                     34586. .34714,35757. .36060,36212. .36315,37016. .37109,
                     37580. .37812,37836. .37921,38120. .38188,39131. .39252,
                     39319. .39444,39499. .40376))
                     /gene="OSJNBa0058E19.4"
                     /codon_start=1
                     /product="Putative gag-pol polyprotein"
                     /protein_id="AAK13129.1"
                     /db_xref="GI:13129471"
                     /translation="MDSIMSNGTWEVVERPYGCKPVGCKWVFKKKLRPDGTIEKYKAR
                     LVAKGYTQKEGEDFFDTYSPVARLTTIRVLLALAASHGLLIHQMDVKTAFLNGELEEE
                     IYMDQPDGYVLESQEGMVCKLLKSLYGLKQAPKQWHEKFDNTLTSAGFVVNEADKCVY
                     YRCGGGEGVILCLYVDDILIFGTSLNVIEEVKDFLSKSFEMTDLRVADVILNIKLLRG
                     DEGGITLVQSHYVDKVLSRFGYSDCKAAPTPYDPSVLLRKNRRIARDQLRYSQIIGSL
                     MYLASATRPDISFAVSKVMRYLKGTMSYGIHYAGYPKVLEGYSDSNWISDADEIKATS
                     GSTMEAELTALDTATVEAEWLRELLMDLPVVEKPVPAILMNCGARRGEARQTAATAAL
                     SRFVTDATRRTEEVLHDSCTSSSSSSVPELQNPSLVYGTSRCLRNSVRLRTCTGRRAI
                     KFLGSASSATAPVLRDYYYCIDAFVNMSMEIGDKEVNDINGGNTASKSSGGTFLGLNF
                     REILLSGRKFTWANSMPSPTYKKLDRVLEFTEEEVKHVVFQMEHKKAPGLDGFPAELI
                     EHKGNFKVFTKVMANRIALVAQKVIRPSQTAFFKGRNIMEGAIIHHETLHEMCKKKSD
                     GVILKLYFEKSYDKENWNFLQQALRMKSFYIWFKWIDQIVSGGSVAVKPKVCIGPDPV
                     LGPGLTLVLVCSTLLIKRGVETYKIYILRERAKPAVACSAASASARAPPSLPRVQPSA
                     AAAPVTAVAPEGHRQVRKVATVLIPSSVSAKDCRSAAAIDDPKTAAASTSLPSSSKPP
                     SFSLLCILACRFRRFEGGDYDPGKVRCFELSYVRGLVTTTQMVTCSAWAHDDMVTYCG
                     VVSCGYSGSPTRLESKRCGQDASVESLRTGGALDTRRTCPAPPNLAKQVEKHPWRYFR
                     QVGSHPAHHAPRGEHTSEDPTISDGAQEPGVIRITRLKGFLQDCVWEPASSDKTADKT
                     RLRVDCLKWPPLYLPDLGRTMNTNIGVSALNRSMTGRMTDGTQDVRRTIACVRRRAAR
                     HIPSNNDGSFERWRDKRRTANAGEGAASSGVPRHPRQGARPPASSPDKAGGQRLVLVP
                     LAGYGARPGSTNRPFKLQEISSRPKVWAKLPSPQRLGNREESGKLRLASTQTTWERFW
                     SGLTPGATVGNMGPGVPTYPHLWRPERSRYAAFNAVCSALNRSMTGRMTDGTQDVRRT
                     IACVRRRAARHIPSNNDGERRSIVGRNLEAWYELVSKVVNTALTDQKDRFFWKKNKDG
                     VVPRKNFIWKLKIPLKIKSFLWYLKERVIFMGTLWARTWPLLLKGVDEEMVKSYCKLL
                     EKRALEFFSMYVVTTMHPSIEEAFHRAPSMHASRIVVTTYKQHQEAEGCEKVKEEEPS
                     MQCGCRRTKFASKRLLSIQSSQRRKWLGFLVWPTTRSCEDHEWSAEVVAWGRCFAAVV
                     LGRAGEPPVGEWMMAVMAPKRIDAYCSIGVGRTG"
     repeat_region   25681. .29265
                     /note="retro02 retroelement"
     repeat_region   30888. .31407
                     /note="RIRE4 Oryza sativa gypsy-type retrotransposon"
     repeat_region   complement(30963. .32572)
                     /note="RIRE8 Oryza sativa gypsy-type retrotransposon"
     repeat_region   31529. .31692
                     /note="Truncator transposon from Oryza sativa"
     repeat_region   complement(31804. .32083)
                     /note="RIRE8 Oryza sativa gypsy-type retrotransposon"
     repeat_region   32562. .34237
                     /note="Truncator transposon"
     repeat_region   complement(32638. .33296)
                     /note="RIRE8 Oryza sativa gypsy-type retrotransposon"
     gene            complement(34475. .43713)
                     /gene="OSJNBa0058E19.5"
                     /note="Putative gag-pol polyprotein"
     CDS             complement(join(34475. .34662,35774. .35971,37596. .37812,
                     39105. .39266,39507. .40550,40983. .42461,42516. .42653,
                     42714. .43024,43070. .43269,43327. .43469,43702. .43713))
                     /gene="OSJNBa0058E19.5"
                     /codon_start=1
                     /product="Putative gag-pol polyprotein"
                     /protein_id="AAK13130.1"
                     /db_xref="GI:13129472"

  Query Match             68.0%;  Score 17;  DB 8;  Length 147706;
  Best Local Similarity   100.0%;  Pred. No. 3.2;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattatt 24
           |||||||||||||||||
```

```
Db  10686 CAAGTCACAGAATTATT 10670


RESULT  22
AP003878
LOCUS       AP003878   150613 bp    DNA             HTG       10-JUL-2001
DEFINITION  Oryza sativa chromosome 8 clone OJ1120_C08, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AP003878
VERSION     AP003878.1  GI:14646811
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1120_C08.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 150613)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, BAC
            clone:OJ1120_C08
  JOURNAL   Published Only in Database (2001) In press
REFERENCE   2  (bases 1 to 150613)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .150613
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="8"
                     /clone="OJ1120_C08"
BASE COUNT    45837 a  30197 c  29818 g  44602 t    159 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 150613;
  Best Local Similarity   100.0%;  Pred. No. 3.2;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattatt 24
          |||||||||||||||||
Db  72720 CAAGTCACAGAATTATT 72736


RESULT  23
AC021110
LOCUS       AC021110   155172 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-238P13, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC021110
VERSION     AC021110.3  GI:8439978
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155172)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155172)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 10, 2000 this sequence version replaced gi:7021826.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0238P13
            ---------- Summary Statistics ----------
            Sequencing vector: M13; 73%
            Sequencing vector: plasmid; 27%
            Chemistry: Dye-primer ET; 73% of reads
            Chemistry: Dye-terminator Big Dye; 27% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 151162 bases at least Q40
            Consensus quality: 152443 bases at least Q30
            Consensus quality: 153249 bases at least Q20
            Insert size: 177000; agarose-fp
            Insert size: 154672; sum-of-contigs
            Quality coverage: 4.08 in Q20 bases; agarose-fp
            Quality coverage: 4.69 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9265: contig of 9265 bp in length
            *    9266     9365: gap of unknown length
            *    9366    26504: contig of 17139 bp in length
            *   26505    26604: gap of unknown length
            *   26605    44300: contig of 17696 bp in length
            *   44301    44400: gap of unknown length
            *   44401    62250: contig of 17850 bp in length
            *   62251    62350: gap of unknown length
            *   62351   105744: contig of 43394 bp in length
            *  105745   105844: gap of unknown length
            *  105845   155172: contig of 49328 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155172
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /clone="RP11-238P13"
     misc_feature    1. .9265
                     /note="assembly_name:Contig1"
     misc_feature    9366. .26504
                     /note="assembly_name:Contig2"
     misc_feature    26605. .44300
                     /note="assembly_name:Contig3
                     clone_end:T7
                     vector_side:right"
     misc_feature    44401. .62250
                     /note="assembly_name:Contig4"
     misc_feature    62351. .105744
                     /note="assembly_name:Contig5"
     misc_feature    105845. .155172
                     /note="assembly_name:Contig6"
BASE COUNT    51543 a  28385 c  28043 g  46698 t    503 others
ORIGIN
```

```
  Query Match             68.0%;  Score 17;  DB 2;  Length 155172;
  Best Local Similarity   100.0%;  Pred. No. 3.2;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaattat 23
          |||||||||||||||||
Db 111929 GCAAGTCACAGAATTAT 111945


RESULT  24
AC092548/c
LOCUS       AC092548   172704 bp    DNA             HTG       14-AUG-2001
DEFINITION  Oryza sativa chromosome 10 clone OSJNBa0073L01, *** SEQUENCING IN
            PROGRESS ***, 10 unordered pieces.
ACCESSION   AC092548
VERSION     AC092548.2  GI:15150615
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 172704)
  AUTHORS   Buell,R., Hsiao,J., Zismann,V., Moffat,K.M., Hill,J.,
            Gansberger,K., Burgess,S., Jarrahi,B., Shvartsbeyn,M., Brenner,M.,
            Ciecko,A., Pai,G., Vanaken,S., Hansen,C., Utterbach,T.,
            Feldblyum,T., Khalak,H.G., Yuan,Q., Quackenbush,J., White,O.,
            Salzberg,S. and Fraser,C.
  TITLE     Oryza sativa ssp. japonica cv. Nipponbare OSJNBa0073L01 BAC genomic
            sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172704)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Aug 14, 2001 this sequence version replaced gi:14787122.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4727: contig of 4727 bp in length
            *     4728     4772: gap of unknown length
            *     4773    20212: contig of 15440 bp in length
            *    20213    20257: gap of unknown length
            *    20258    87269: contig of 67012 bp in length
            *    87270    87314: gap of unknown length
            *    87315    96752: contig of 9438 bp in length
            *    96753    96797: gap of unknown length
            *    96798   106410: contig of 9613 bp in length
            *   106411   106454: gap of unknown length
            *   106455   117131: contig of 10677 bp in length
            *   117132   117175: gap of unknown length
            *   117176   129369: contig of 12194 bp in length
            *   129370   129413: gap of unknown length
            *   129414   143358: contig of 13945 bp in length
            *   143359   143402: gap of unknown length
            *   143403   157616: contig of 14214 bp in length
            *   157617   157660: gap of unknown length
            *   157661   172704: contig of 15044 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172704
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:4530"
                     /chromosome="10"
                     /clone="OSJNBa0073L01"
BASE COUNT    49372 a   35151 c   36535 g   51240 t    406 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 172704;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattatt 24
          |||||||||||||||||
Db  95655 CAAGTCACAGAATTATT 95639


RESULT  25
AP001390/c
LOCUS       AP001390   174815 bp    DNA             HTG       30-MAY-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-635B11 map 18q23, WORKING
            DRAFT SEQUENCE, 21 unordered pieces.
ACCESSION   AP001390
VERSION     AP001390.2  GI:8117301
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-635B11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174815)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 174,815 genomic DNA of 18q23
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 174815)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7259325.
            ---------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            ---------- Project Information
                Center project name: HumDraft18
                Center clone name: RP11-635B11
            ---------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 159194 bases at least Q40
                Consensus quality: 167087 bases at least Q30
                Consensus quality: 170693 bases at least Q20
                Insert size: 172815; sum-of-contigs
                Quality coverage: 4.94x in Q20 bases; sum-of-contigs
            ----------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            21 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     58823 contig of  58823 bp in length
              58924     73243 contig of  14320 bp in length
              73344     83572 contig of  10229 bp in length
              83673     94547 contig of  10875 bp in length
              94648    104597 contig of   9950 bp in length
             104698    114347 contig of   9650 bp in length
```

```
            114448    122080 contig of   7633 bp in length
            122181    128465 contig of   6285 bp in length
            128566    135791 contig of   7226 bp in length
            135892    142934 contig of   7043 bp in length
            143035    147999 contig of   4965 bp in length
            148100    151734 contig of   3635 bp in length
            151835    155576 contig of   3742 bp in length
            155677    159193 contig of   3517 bp in length
            159294    161359 contig of   2066 bp in length
            161460    163515 contig of   2056 bp in length
            163616    166614 contig of   2999 bp in length
            166715    169056 contig of   2342 bp in length
            169157    171527 contig of   2371 bp in length
            171628    173248 contig of   1621 bp in length
            173349    174815 contig of   1467 bp in length
        Sequence updated (26-May-2000).
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 21 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *        1     58823: contig of 58823 bp in length
        *    58824 58923: gap of      100 bp
        *    58924     73243: contig of 14320 bp in length
        *    73244 73343: gap of      100 bp
        *    73344     83572: contig of 10229 bp in length
        *    83573 83672: gap of      100 bp
        *    83673     94547: contig of 10875 bp in length
        *    94548 94647: gap of      100 bp
        *    94648    104597: contig of 9950 bp in length
        *   104598 104697: gap of      100 bp
        *   104698    114347: contig of 9650 bp in length
        *   114348 114447: gap of      100 bp
        *   114448    122080: contig of 7633 bp in length
        *   122081 122180: gap of      100 bp
        *   122181    128465: contig of 6285 bp in length
        *   128466 128565: gap of      100 bp
        *   128566    135791: contig of 7226 bp in length
        *   135792 135891: gap of      100 bp
        *   135892    142934: contig of 7043 bp in length
        *   142935 143034: gap of      100 bp
        *   143035    147999: contig of 4965 bp in length
        *   148000 148099: gap of      100 bp
        *   148100    151734: contig of 3635 bp in length
        *   151735 151834: gap of      100 bp
        *   151835    155576: contig of 3742 bp in length
        *   155577 155676: gap of      100 bp
        *   155677    159193: contig of 3517 bp in length
        *   159194 159293: gap of      100 bp
        *   159294    161359: contig of 2066 bp in length
        *   161360 161459: gap of      100 bp
        *   161460    163515: contig of 2056 bp in length
        *   163516 163615: gap of      100 bp
        *   163616    166614: contig of 2999 bp in length
        *   166615 166714: gap of      100 bp
        *   166715    169056: contig of 2342 bp in length
        *   169057 169156: gap of      100 bp
        *   169157    171527: contig of 2371 bp in length
        *   171528 171627: gap of      100 bp
        *   171628    173248: contig of 1621 bp in length
        *   173249 173348: gap of      100 bp
        *   173349    174815: contig of 1467 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174815
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18q23"
                     /clone="RP11-635B11"
     misc_feature    1. .58823
                     /note="assembly_fragment"
     misc_feature    58924. .73243
                     /note="assembly_fragment"
     misc_feature    73344. .83572
                     /note="assembly_fragment"
     misc_feature    83673. .94547
                     /note="assembly_fragment"
     misc_feature    94648. .104597
                     /note="assembly_fragment"
     misc_feature    104698. .114347
                     /note="assembly_fragment"
     misc_feature    114448. .122080
                     /note="assembly_fragment"
     misc_feature    122181. .128465
                     /note="assembly_fragment"
     misc_feature    128566. .135791
                     /note="assembly_fragment"
     misc_feature    135892. .142934
                     /note="assembly_fragment"
     misc_feature    143035. .147999
                     /note="assembly_fragment"
     misc_feature    148100. .151734
                     /note="assembly_fragment"
     misc_feature    151835. .155576
                     /note="assembly_fragment"
     misc_feature    155677. .159193
                     /note="assembly_fragment"
     misc_feature    159294. .161359
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    161460. .163515
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    163616. .166614
                     /note="assembly_fragment"
     misc_feature    166715. .169056
                     /note="assembly_fragment"
     misc_feature    169157. .171527
                     /note="assembly_fragment"
     misc_feature    171628. .173248
                     /note="assembly_fragment"
     misc_feature    173349. .174815
                     /note="assembly_fragment"
BASE COUNT    52310 a  30923 c  32183 g  57399 t   2000 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 174815;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaattat 23
           |||||||||||||||||
Db   44384 GCAAGTCACAGAATTAT 44368


RESULT  26
AC091589/c
LOCUS       AC091589   178056 bp    DNA             HTG       26-JUL-2001
DEFINITION  Homo sapiens chromosome 18 clone RP11-635B11 map 18, *** SEQUENCING
            IN PROGRESS ***, 2 ordered pieces.
ACCESSION   AC091589
VERSION     AC091589.6  GI:15022066
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178056)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-635B11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178056)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
```

```
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 26, 2001 this sequence version replaced gi:14994291.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L13182
                Center clone name: 635_B_11
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   168134: contig of 168134 bp in length
            *  168135 168234: gap of      100 bp
            *  168235   178056: contig of 9822 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178056
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-635B11"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    49585 a  30884 c  34094 g  63393 t    100 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 178056;
  Best Local Similarity   100.0%;  Pred. No. 3.1;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaattat 23
           |||||||||||||||||
Db   64175 GCAAGTCACAGAATTAT 64159


RESULT  27
AC002523/c
LOCUS       AC002523    228516 bp    DNA             PRI       05-MAY-1998
DEFINITION  Homo sapiens Xq28 BACs 360 F12, GSHB-555C13, complete sequence.
ACCESSION   AC002523 AC002374 AC002373
VERSION     AC002523.1  GI:2815518
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 228516)
  AUTHORS   Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,
            Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Haywood,M.,
            Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L., Shen,H.,
            Worley,K., Chen,E., Forcum,J., Arenson,A.D., Gorrell,J.H.,
            Brundage,E., Di,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 228516)
  AUTHORS   Chiu,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 228516)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 228516)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-1998) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT     Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.

            Begining of sequence overlaps with AF007262, end of sequence
            overlaps with AF011889.
FEATURES             Location/Qualifiers
     source          1. .228516
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="360 F12, GSHB-555C13"
                     /chromosome="X"
     misc_feature    1. .149
                     /note="Positions 104645 to 104793 in AF007262 overlap with
                     postions 1-149 in this clone"
                     /note="Region: Overlap with AF007262"
     repeat_region   complement(487. .547)
                     /rpt_family="MER5B"
     repeat_region   complement(1270. .1577)
                     /rpt_family="MLT1E"
     repeat_region   complement(1653. .1708)
                     /rpt_family="MLT1D"
     repeat_region   complement(2282. .2306)
                     /rpt_family="AT_rich"
     repeat_region   3316. .3617
                     /rpt_family="MLT1B"
     repeat_region   3618. .3651
                     /rpt_family="(CA)n"
     repeat_region   complement(3658. .3944)
                     /rpt_family="AluSx"
     repeat_region   3949. .4012
                     /rpt_family="MLT1B"
```

```
repeat_region   complement(4098. .4386)
                /rpt_family="LINE2"
repeat_region   4723. .4784
                /rpt_family="MIR"
repeat_region   5630. .5727
                /rpt_family="(GA)n"
repeat_region   7337. .7584
                /rpt_family="LINE2"
repeat_region   7678. .7709
                /rpt_family="(CA)n"
repeat_region   7710. .7863
                /rpt_family="LINE2"
repeat_region   complement(7981. .8151)
                /rpt_family="LINE2"
repeat_region   complement(8177. .8274)
                /rpt_family="LINE2"
repeat_region   complement(10118. .10192)
                /rpt_family="LINE2"
repeat_region   complement(10400. .10443)
                /rpt_family="AT_rich"
repeat_region   complement(10668. .10695)
                /rpt_family="AT_rich"
repeat_region   11811. .11896
                /rpt_family="LINE2"
repeat_region   12526. .12605
                /rpt_family="LINE2"
repeat_region   14892. .14932
                /rpt_family="MLT1A2"
repeat_region   15361. .15661
                /rpt_family="AluSp"
repeat_region   16706. .16765
                /rpt_family="AT_rich"
repeat_region   17193. .17764
                /rpt_family="L1PA7"
repeat_region   17811. .18912
                /rpt_family="L1PA7"
repeat_region   complement(18968. .19018)
                /rpt_family="L1"
repeat_region   19034. .19085
                /rpt_family="L1"
repeat_region   complement(19084. .19727)
                /rpt_family="L1"
repeat_region   19737. .19971
                /rpt_family="L1"
repeat_region   19971. .20406
                /rpt_family="L1ME3A"
repeat_region   20452. .20500
                /rpt_family="AT_rich"
repeat_region   20899. .21008
                /rpt_family="(CAAAA)n"
repeat_region   21061. .21098
                /rpt_family="(TA)n"
repeat_region   complement(21283. .21343)
                /rpt_family="AT_rich"
repeat_region   complement(21527. .22106)
                /rpt_family="MLT1F"
repeat_region   complement(22286. .22326)
                /rpt_family="AT_rich"
repeat_region   23171. .23316
                /rpt_family="MER44A"
repeat_region   23311. .23626
                /rpt_family="MER44C"
repeat_region   complement(23681. .23757)
                /rpt_family="AT_rich"
repeat_region   24613. .24924
                /rpt_family="AluSx"
repeat_region   complement(24927. .25278)
                /rpt_family="MSTA"
repeat_region   25554. .25668
                /rpt_family="MIR"
repeat_region   complement(27230. .27359)
                /rpt_family="MER45"
repeat_region   complement(28025. .28081)
                /rpt_family="MER45"
repeat_region   28227. .28312
                /rpt_family="MIR"
repeat_region   complement(28521. .28607)
                /rpt_family="MIR"
repeat_region   complement(29882. .29953)
                /rpt_family="AT_rich"
repeat_region   29960. .30154
                /rpt_family="MIR"
repeat_region   complement(30336. .30383)
                /rpt_family="AT_rich"
repeat_region   complement(31519. .31757)
                /rpt_family="MIR"
repeat_region   32198. .32905
                /rpt_family="L1MA9"
repeat_region   33664. .33841
                /rpt_family="MIR"
repeat_region   complement(34528. .34559)
                /rpt_family="AT_rich"
repeat_region   complement(34576. .34659)
                /rpt_family="MIR"
repeat_region   complement(34735. .35097)
                /rpt_family="MSTD"
repeat_region   complement(35249. .35546)
                /rpt_family="AluY"
repeat_region   complement(36693. .36767)
                /rpt_family="(CA)n"
repeat_region   complement(37299. .37786)
                /rpt_family="MLT1D"
STS             38010. .38096
                /note="L47526, HUMSWX2596, Chr. X"
                /db_xref="dbSTS: 17731"
repeat_region   38232. .38255
                /rpt_family="AT_rich"
repeat_region   39138. .39218
                /rpt_family="MIR"
repeat_region   complement(39901. .40077)
                /rpt_family="L1MC3"
repeat_region   complement(40079. .40208)
                /rpt_family="AluSg/x"
repeat_region   complement(40209. .40353)
                /rpt_family="L1MC3"
repeat_region   complement(40354. .40646)
                /rpt_family="MER47B"
repeat_region   complement(40647. .40798)
                /rpt_family="L1MC3"
repeat_region   complement(40793. .42299)
                /rpt_family="L1MC3"
repeat_region   complement(42297. .42391)
                /rpt_family="L1MD3"
repeat_region   complement(42406. .45376)

  Query Match            68.0%;  Score 17;  DB 9;  Length 228516;
  Best Local Similarity  100.0%;  Pred. No. 3.1;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaattattt 25
          |||||||||||||||||
Db  55846 AAGTCACAGAATTATTT 55830


RESULT  28
AP001690/c
LOCUS       AP001690   340000 bp    DNA             PRI       30-MAY-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q, section 34/105.
ACCESSION   AP001690 AL163235 BA000005
VERSION     AP001690.1  GI:7768701
KEYWORDS    .
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
```

```
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.
  TITLE     The DNA sequence of human chromosome 21. The chromosome 21 mapping
            and sequencing consortium
  JOURNAL   Nature 405 (6784), 311-319 (2000)
  MEDLINE   20289799
REFERENCE   2  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The
            Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic
            Sciences Center, Human Genome Research Group * Institute of
            Molecular Biotechnology, Genome Analysis * Keio University School
            of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome
            Analysis * Max-Planck Institute for Molecular Genetics (addresses
            see below)
COMMENT     On May 30, 2000 this sequence version replaced gi:7717288.
            The chromosome 21 mapping and sequencing consortium consisting of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagamihara 228-8555, Japan,
            * e.mail: hattori@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Molecular Biology, * Tokyo
            160-8582, Japan,
            * e.mail: nshimizu@dmb-med.keio.ac.jp
            * URL: http://www.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/
            AL163235: Submitted (10-Apr-2000).
FEATURES             Location/Qualifiers
     source          1. .340000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
     source          <1. .114687
                     /note="Accession No. AP000469"
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21.2"
                     /clone="R132H24, 5' partial"
                     /clone_lib="RPCI-11 BAC library"
     repeat_region   complement(1162. .1526)
                     /note="L1PA13"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   1527. .1893
                     /note="L1PA13"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   1905. .2287
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   2386. .2549
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   complement(4287. .4410)
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   complement(4489. .4670)
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   complement(6498. .6745)
                     /note="L1MA9"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   7600. .7882
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(7912. .8041)
                     /note="AluJb"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(8745. .9054)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(9504. .9794)
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   9983. .10564
                     /note="L1MB8"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   10720. .10787
                     /note="(TAGA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   11137. .11436
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     STS             13135. .13345
                     /standard_name="D21S1914"
                     /note="AFM344wf5
                     Accession No. Z51475"
     repeat_region   13259. .13313
                     /note="(TG)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   complement(14329. .14632)
                     /note="L1MA3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   14625. .14852
                     /note="L1MA3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   14853. .15154
                     /note="AluY"
                     /rpt_family="SINE/Alu"
```

```
                     /rpt_type=DISPERSED
     repeat_region   15155. .15634
                     /note="L1MA3"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   complement(15694. .15808)
                     /note="L1MA6"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   15840. .16145
                     /note="AluY"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   17627. .18220
                     /note="L1"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   18201. .18826
                     /note="L1"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   18827. .19126
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   19127. .19352
                     /note="L1"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   19353. .20555
                     /note="L1PA7"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   20557. .21024
                     /note="L1P5"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   21026. .21055
                     /note="(TG)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   complement(21414. .21879)
                     /note="L1MA5A"
                     /rpt_family="LINE/L1"

  Query Match             68.0%;  Score 17;  DB 9;  Length 340000;
  Best Local Similarity   100.0%;  Pred. No. 2.9;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattattt 25
           |||||||||||||||||
Db  141812 AAGTCACAGAATTATTT 141796


RESULT  29
HSAPAF20/c
LOCUS       HSAPAF20      605 bp    DNA             PRI       03-AUG-2000
DEFINITION  Homo sapiens APAF1 gene, exon 12, 3' end.
ACCESSION   AF098889
VERSION     AF098889.1  GI:4191520
KEYWORDS    .
SEGMENT     20 of 47
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 605)
  AUTHORS   Bala,S., Oliver,H., Renault,B., Montgomery,K., Dutta,S., Rao,P.,
            Houldsworth,J., Kucherlapati,R., Wang,X., Chaganti,R.S. and
            Murty,V.V.
  TITLE     Genetic analysis of the APAF1 gene in male germ cell tumors
  JOURNAL   Genes Chromosomes Cancer 28 (3), 258-268 (2000)
  MEDLINE   20320642
```

```
    PUBMED   10862031
REFERENCE   2  (bases 1 to 605)
  AUTHORS   Bala,S., Oliver,H., Renault,B., Dutta,S., Montgomery,K.,
            Kucherlapati,R., Wang,X., Chaganti,R.S.K. and Murty,V.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1998) Pathology, College of Physicians & Surgeons
            of Columbia University, 630 West 168th Street, New York, New York
            10032, USA
FEATURES             Location/Qualifiers
     source          1. .605
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     exon            <1. .70
                     /gene="APAF1"
                     /number=12
     intron          71. .>605
                     /gene="APAF1"
                     /number=12
BASE COUNT      184 a     93 c    134 g    192 t      2 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 9;  Length 605;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattatt 24
           ||||||||||||||||
Db     425 AAGTCACAGAATTATT 410


RESULT  30
G07181/c
LOCUS       G07181       1829 bp    DNA             STS       19-OCT-1995
DEFINITION  human STS WI-9211.
ACCESSION   G07181
VERSION     G07181.1  GI:860426
KEYWORDS    STS sequence; primer; sequence tagged site.
SOURCE      human STSs derived from sequences in dbEST and the Unigene
            collection.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
            Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1829)
  AUTHORS   Hudson,T.
  TITLE     Whitehead Institute/MIT Center for Genome Research; Physically
            Mapped ESTs
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Thomas Hudson
            Whitehead Institute/MIT Center for Genome Research
            Whitehead Institute for Biomedical Research
            9 Cambridge Center, Cambridge MA 02142 USA
            Tel: 617 252 1900
            Fax: 617 252 1902
            Email: thudson@genome.wi.mit.edu

            Primer A: CAGCCACACTAACTGCATTAGC
            Primer B: GACAGAAATGGATTGTATAAATGGC
            STS size: 111
            PCR Profile:
                  Presoak:
                  Denaturation:
                  Annealing: 56 degrees C
                  Polymerization:
                  PCR Cycles: 35
                  Thermal Cycler:
            Protocol:
                Template: 10 ng
                Primer: each 5 pM
                dNTPs: each 4 nM
```

```
            Taq Polymerase: 0.025 units/ul
            Total Vol: 20 ul

          Buffer:
            MgCl2: 1.5 mM
            KCl: 50 mM
            Tris-HCL: 10 mM
            pH: 9.3

          Prepared with primer pairs derived from D13644 -- Unigene.
FEATURES             Location/Qualifiers
     source          1. .1829
                     /organism="Homo sapiens"
     STS             891. .1001
     primer_bind     891. .912
     primer_bind     complement(977. .1001)
BASE COUNT      539 a    321 c    320 g    590 t     59 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 1829;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db   1689 CAAGTCACAGAATTAT 1674


RESULT  31
D13644/c
LOCUS       D13644       4602 bp    mRNA            PRI       14-JUN-2001
DEFINITION  Homo sapiens mRNA for KIAA0019 protein, complete cds.
ACCESSION   D13644
VERSION     D13644.1  GI:1531551
KEYWORDS    KIAA0019.
SOURCE      Homo sapiens male myoblast cell_line:KG-1 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y.,
            Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
  TITLE     Prediction of the coding sequences of unidentified human genes. I.
            The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
            analysis of randomly sampled cDNA clones from human immature
            myeloid cell line KG-1
  JOURNAL   DNA Res. 1 (1), 27-35 (1994)
  MEDLINE   96051387
REFERENCE   2  (sites)
  AUTHORS   Nomura,N., Miyajima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y.,
            Sato,S., Nagase,T., Seki,N., Ishikawa,K. and Tabata,S.
  TITLE     Prediction of the coding sequences of unidentified human genes. I.
            The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
            analysis of randomly sampled cDNA clones from human immature
            myeloid cell line KG-1 (supplement)
  JOURNAL   DNA Res. 1 (1), 47-56 (1994)
  MEDLINE   96051389
REFERENCE   3  (bases 1 to 4602)
  AUTHORS   Matoskova,B., Wong,W.T., Seki,N., Nagase,T., Nomura,N.,
            Robbins,K.C. and Di Fiore,P.P.
  TITLE     RN-tre identifies a family of tre-related proteins displaying a
            novel potential protein binding domain
  JOURNAL   Oncogene 12 (12), 2563-2571 (1996)
  MEDLINE   96293402
REFERENCE   4  (bases 1 to 4602)
  AUTHORS   Nomura,N., Miyajima,N., Kawarabayasi,Y. and Tabata,S.
  TITLE     Prediction of new human genes by entire sequencing of randomly
            sampled cDNA clones
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 4602)
  AUTHORS   Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-1992) Nobuo Nomura, Kazusa DNA Research
            Institute, Gene Structure 1; 1532-3 Yana, Kisarazu, Chiba 292,
            Japan (E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp,
            Tel:0438-52-3930, Fax:0438-52-3931)
COMMENT     On Sep 10, 1996 this sequence version replaced gi:286016.
FEATURES             Location/Qualifiers
     source          1. .4602
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /map="10p13"
                     /sex="male"
                     /cell_line="KG-1"
                     /cell_type="myoblast"
     5'UTR           <1. .279
     gene            280. .2766
                     /gene="KIAA0019"
     CDS             280. .2766
                     /gene="KIAA0019"
                     /note="protein related N-ternimus of tre oncogene"
                     /codon_start=1
                     /product="KIAA0019 protein"
                     /protein_id="BAA02807.1"
                     /db_xref="GI:2104571"
                     /translation="MNSDQDVALKLAQERAEIVAKYDRGREGAEIEPWEDADYLVYKV
                     TDRFGFLHEEELPDHNVAVERQKHLEIERTTKWLKMLKGWEKYKNTEKFHRRIYKGIP
                     LQLRGEVWALLLEIPKMKEETRDLYSKLKHRARGCSPDIRQIDLDVNRTFRDHIMFRD
                     RYGVKQQSLFHVLAAYSIYNTEVGYCQGMSQITALLLMYMNEEDAFWALVKLFSGPKH
                     AMHGFFVQGFPKLLRFQEHHEKILNKFLSKLKQHLDSQEIYTSFYTMKWFFQCFLDRT
                     PFTLNLRIWDIYIFEGERVLTAMSYTILKLHKKHLMKLSMEELVEFFQETLAKDFFFE
                     DDFVIEQLQISMTELKRAKLDLPEPGKEDEYPKKPLGQLPPELQSWGVHHLSNGQRSV
                     GRPSPLASGRRESGAPHRRHEHSPHPQSRTGTPERAQPPRRKSVEEESKKLKDEADFQ
                     RKLPSGPQDSSRQYNHAAANQNSNATSNIRKEFVPKWNKPSDVSATERTAKYTMEGKG
                     RAAHPALAVTVPGPAEVRVSNVRPKMKALDAEDGKRGSTASQYDNVPGPELDSGASVE
                     EALERAYSQSPRHALYPPSPRKHAEPSSSPSKVSNKFTFKVQPPSHARYPSQLDGEAR
                     GLAHPPSYSNPPVYHGNSPKHFPTANSSFASPQFSPGTQLNPSRRPHGSTLSVSASPE
                     KSYSRPSPLVLPSSRIEVLPVDTGAGGYSGNSGSPKNGKLIIPPVDYLPDNRTWSEVS
                     YTYRPETQGQSWTRDASRGNLPKYSAFQLAPFQDHGLPAVSVDSPVRYKASPAAEDAS
                     PSGYPYSGPPPPAYHYRNRDGLSIQESVLL"
     3'UTR           2767. .4602
BASE COUNT     1340 a   1041 c    999 g   1222 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 9;  Length 4602;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db   4462 CAAGTCACAGAATTAT 4447


RESULT  32
LLA012388
LOCUS       LLA012388    5469 bp    DNA             BCT       29-OCT-1999
DEFINITION  Lactococcus lactis dar gene and genes encoding five hypothetical
            proteins.
ACCESSION   AJ012388
VERSION     AJ012388.1  GI:6165401
KEYWORDS    acetoin(diacetyl)reductase; dar gene; ORF1; ORF3; ORF4; ORF5; ORF6;
            putative.
SOURCE      Lactococcus lactis.
  ORGANISM  Lactococcus lactis
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Lactococcus.
REFERENCE   1  (bases 1 to 5469)
  AUTHORS   Aungpraphapornchai,P., Griffin,H.G. and Gasson,M.J.
  TITLE     Cloning, DNA sequence analysis, and deletion of a gene encoding
            diacetyl-acetoin reductase from Lactococcus lactis
  JOURNAL   DNA Seq. 10 (3), 163-172 (1999)
  MEDLINE   20112344
REFERENCE   2  (bases 1 to 5469)
```

```
  AUTHORS   Aungpraphapornchai,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-OCT-1998) Aungpraphapornchai P., Genetics and
            Microbiology, Institute of Food Research, Colney lane, Norwich,
            Norfolk, NR4 7UA, UK
FEATURES             Location/Qualifiers
     source          1. .5469
                     /organism="Lactococcus lactis"
                     /sub_species="cremoris"
                     /strain="MG1363"
                     /db_xref="taxon:1358"
     RBS             77. .80
                     /note="ORF1"
     CDS             87. .950
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB59825.1"
                     /db_xref="GI:6165402"
                     /translation="MNPKNRNIIIAVAVLILVALVAFFSLNHQGGVKASAGEKTVKVG
                     IMSGDKQDQEVWKSVANTAKEKYDLKLKFVYFSDYNQPNEALLSGDIDINAFQSYNYV
                     KTWNKAHKSDIVAVGNTYITPMHIYSKEISKLSDLKEGSTVAIPNDASNESRALFVLQ
                     SAGLLKLTTSDSSKLVGLPDITENPHQLKFKEVDASQTPRALDSVALSVVNYNYATAA
                     SLPKSESVFMEPLNKTSAQYINFIATTSKEKNNKVYKEVAKAYASKATEKAIKEQYPD
                     GGELPAWDLKL"
     RBS             1005. .1009
                     /gene="dar"
     gene            1005. .1789
                     /gene="dar"
     CDS             1019. .1789
                     /gene="dar"
                     /codon_start=1
                     /transl_table=11
                     /product="diacetyl(acetoin) reductase"
                     /protein_id="CAB59826.1"
                     /db_xref="GI:6165403"
                     /translation="MGKLAIVTGSGQGIGEGIVYRLAKDGFAVAVADINPQTAKNVAK
                     TLIDKGFEAKAYVLDVANRSAVFQLIEQAVSDFGELAVFVNNAGVAFIDSLIDSDSER
                     VERLFDVNIKGTYWGIQAAATQFKKQGHGGRIINAASLASVEGSALQSAYSASKFAIR
                     GLTQSAAKELAKDYITVNAYNPGIVRTAMRDEIDKKTAQIKGISIEQQQQNCLNEIYL
                     GREGRPEDVAEVVAWFASDAAEYITGQSILVDGGMRFN"
     RBS             1794. .1798
                     /note="ORF3"
     CDS             1805. .2665
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB59827.1"
                     /db_xref="GI:6165404"
                     /translation="MNVKNRRILITIIILVFIIIVGGIFAFSHSGNKSKVSSKIVKIG
                     LMPGGKQEDVIWKQVQKNAKDQFGITLKFVNFTDGDEPNKALVNHEVDLNAFQHYAYL
                     KSWNKANNGNIVSIGDTIITPIHLYSTKYKKVDEIPDKSTIAIPNDITNESRALYVLK
                     NAGLIKLDTSRGVLATVKDIRENPKSLIIKEIDASQTPRALDSVAAAVINYNFAISAK
                     NSDKESIYQEPLNEDSAQWINFIAANQSDKNNKVYKEVVKAYEQKNIADIIKKEYPDG
                     GELPAWNLKL"
     RBS             2776. .2781
                     /note="ORF4"
     CDS             2789. .3895
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAB59828.1"
                     /db_xref="GI:6165405"
                     /translation="MTAIIELNNLSVQFHQKGRLVTAVKNATLHIEKGDIYGVIGYSG
                     AGKSTLVRTINLLQKPTEGQIVINGEKIFDSENPVKFTGAKLREFRQKIGMIFQHFNL
                     LSEKTVFNNVAFALQHSQIEDKNGKKRYLTKKEKNDKVTELLKLVDLADLSDKYPAQL
                     SGGQKQRVAIARALANDPEILISDEGTSALDPKTTNQILDLLKSLHEKLGITVVLITH
                     EMQVVKEIANKVAVMQNGEIIEQNSLIDIFAQPKEALTKQFIETTSSVNRFIASLSKT
                     ELLAQLADDEELIHLDYSGSELEDPVVSDITKKFDVTTNIFYGNVELLQGQPFGSLVL
                     TLKGSSEHRAAAKAYFVERHLKFEVLGKIERTVD"
```

```
          RBS             3878. .3884
                          /note="ORF5"
          CDS             3895. .4590
                          /note="ORF5"
                          /codon_start=1
                          /transl_table=11
                          /product="hypothetical protein"
                          /protein_id="CAB59829.1"
                          /db_xref="GI:6165406"
                          /translation="MAEWFAHTFPNVVYLGWTGETGWWTAIVQTLYMTFISALIGGLL
                          GLIFGIGVVVTAEDGITPNRPLFWILDKIVSIGRAFPFIILLAAIAPLTKILVGTQIG
                          VTAALVPLALGVAPFYARQVQASLESVDHGKVEAAQTVGADFLDIVFTVYLREELASL
                          IRVSTVTLISLIGLTAMAGAIGAGGLGNTAISYGYNRFANDVTWFATILILIFVLLVQ
                          LVGDFLARRVSHR"
          RBS             4621. .4627
                          /note="ORF6"
          CDS             4633. .5181
                          /note="ORF6"
                          /codon_start=1
                          /transl_table=11
                          /product="hypothetical protein"
                          /protein_id="CAB59830.1"
                          /db_xref="GI:6165407"
                          /translation="MKNNSVKIVVATGIGAALFVIIGWLINIPTPIPNTSIQLQYAVL
                          ALFSALFGPLAGFLIGFIGHALKDSFLYGAPWWTWVLGSGLMGLFLGFGVKRESLTQG
                          IFGNKEIIRFNIVQFLANVVVWGLIAPIGDILVYSEPANKVFTQGVVAGLVNALTIAV
                          AGTLLLKLYAATRTKSGTLDKE"
          terminator      5210. .5235
                          /note="putative"
     BASE COUNT     1796 a    858 c   1134 g   1681 t
     ORIGIN


       Query Match             64.0%;  Score 16;  DB 1;  Length 5469;
       Best Local Similarity   100.0%;  Pred. No. 18;
       Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

     Qy       9 aagtcacagaattatt 24
                ||||||||||||||||
     Db    3204 AAGTCACAGAATTATT 3219


     RESULT  33
     AL161800/c
     LOCUS       AL161800    43741 bp    DNA             PRI       04-APR-2001
     DEFINITION  Human DNA sequence from clone RP11-318C17 on chromosome 20.
                 Contains part of a novel gene and GSSs, complete sequence.
     ACCESSION   AL161800
     VERSION     AL161800.10  GI:8246908
     KEYWORDS    HTG.
     SOURCE      human.
       ORGANISM  Homo sapiens
                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
     REFERENCE   1  (bases 1 to 43741)
       AUTHORS   Wilson,S.
       TITLE     Direct Submission
       JOURNAL   Submitted (15-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
                 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
                 requests: clonerequest@sanger.ac.uk
     COMMENT     On Jun 4, 2000 this sequence version replaced gi:8217673.
                 During sequence assembly data is compared from overlapping clones.
                 Where differences are found these are annotated as variations
                 together with a note of the overlapping clone name. Note that the
                 variation annotation may not be found in the sequence submission
                 corresponding to the overlapping clone, as we submit sequences with
                 only a small overlap as described above.
                 The following abbreviations are used to associate primary accession
                 numbers given in the feature table with their source databases:
                 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
                 on the WORMPEP database can be found at
                 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
                 was generated from part of bacterial clone contigs of human
```

```
          chromosome 20, constructed by the Sanger Centre Chromosome 20
          Mapping Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr20
          IMPORTANT: This sequence is not the entire insert of clone
          RP11-318C17 It may be shorter because we sequence overlapping
          sections only once, except for a 100 base overlap.
          The true left end of clone RP5-974N19 is at 43642 in this sequence.
          The true right end of clone RP5-858M22 is at 100 in this sequence.
          This sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest. RP11-318C17 is from
          the library RPCI-11.2 constructed by the group of Pieter de Jong.
          For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .43741
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP11-318C17"
                     /clone_lib="RPCI-11.2"
     repeat_region   3. .293
                     /note="AluSx repeat: matches 1. .291 of consensus"
     repeat_region   390. .683
                     /note="AluSg repeat: matches 1. .286 of consensus"
     repeat_region   788. .3327
                     /note="L1PA16 repeat: matches 3587. .6157 of consensus"
     misc_feature    3411. .3820
                     /note="match: GSS: Em:AQ371040"
     repeat_region   4157. .4477
                     /note="L2 repeat: matches 2348. .2710 of consensus"
     misc_feature    complement(6523. .6967)
                     /note="match: GSS: Em:AQ236127"
     misc_feature    complement(6607. .6950)
                     /note="match: GSS: Em:AQ110431"
     repeat_region   6676. .6820
                     /note="MIR repeat: matches 79. .229 of consensus"
     repeat_region   7231. .7405
                     /note="MIR repeat: matches 42. .206 of consensus"
     repeat_region   7416. .7551
                     /note="L2 repeat: matches 2584. .2710 of consensus"
     repeat_region   9521. .9605
                     /note="MLT1C repeat: matches 433. .523 of consensus"
     mRNA            complement(join(9542. .9883,13385. .>13593))
                     /gene="bA318C17.1"
                     /note="continued from dJ974N19.2 in Em:AL138808
                     match: ESTs: Em:AL542749 Em:AL569592"
                     /evidence=not_experimental
                     /product="bA318C17.1 (novel transcript)"
     polyA_site      complement(9542)
                     /gene="bA318C17.1"
     gene            complement(join(9542. .9883,13385. .13593))
                     /gene="bA318C17.1"
     polyA_signal    complement(9565. .9570)
                     /gene="bA318C17.1"
     repeat_region   9805. .9884
                     /note="MLT1G repeat: matches 59. .137 of consensus"
     repeat_region   9885. .10320
                     /note="MER4B repeat: matches 1. .437 of consensus"
     repeat_region   10325. .10397
                     /note="MER4C repeat: matches 393. .465 of consensus"
     repeat_region   10682. .11085
                     /note="L2 repeat: matches 1981. .2406 of consensus"
     repeat_region   11167. .11657
                     /note="L2 repeat: matches 2263. .2750 of consensus"
     repeat_region   11872. .12002
                     /note="L2 repeat: matches 2621. .2750 of consensus"
     repeat_region   12681. .12976
                     /note="AluJb repeat: matches 1. .305 of consensus"
     repeat_region   14480. .15774
                     /note="L2 repeat: matches 1424. .2749 of consensus"
     repeat_region   15775. .16067
                     /note="AluSx repeat: matches 1. .310 of consensus"
     repeat_region   16068. .16216
                     /note="L2 repeat: matches 1285. .1424 of consensus"
     repeat_region   16713. .16865
                     /note="L2 repeat: matches 2545. .2699 of consensus"
     repeat_region   17484. .17764
                     /note="AluSx repeat: matches 6. .296 of consensus"
     repeat_region   17813. .18013
                     /note="L1MB2 repeat: matches 5966. .6164 of consensus"
     repeat_region   18014. .18267
                     /note="AluSc repeat: matches 1. .252 of consensus"
     repeat_region   18268. .18773
                     /note="L1MB2 repeat: matches 5450. .5966 of consensus"
     repeat_region   19067. .19374
                     /note="AluSg1 repeat: matches 1. .304 of consensus"
     repeat_region   19513. .19811
                     /note="AluYa8 repeat: matches 1. .305 of consensus"
     repeat_region   21304. .21613
                     /note="AluSg repeat: matches 1. .308 of consensus"
     repeat_region   22998. .23343
                     /note="MLT1A1 repeat: matches 1. .353 of consensus"
     repeat_region   23736. .23865
                     /note="AluSg repeat: matches 1. .130 of consensus"
     repeat_region   24570. .25025
                     /note="MLT1C repeat: matches 1. .466 of consensus"
     repeat_region   25235. .26166
                     /note="L1M4 repeat: matches 3073. .4010 of consensus"
     misc_feature    complement(26577. .27248)
                     /note="match: GSS: Em:AQ036701"
     repeat_region   28315. .28629
                     /note="AluSq repeat: matches 1. .313 of consensus"
     repeat_region   28933. .29096
                     /note="MLT1I repeat: matches 245. .409 of consensus"
     repeat_region   29893. .30110
                     /note="MIR repeat: matches 25. .245 of consensus"
     misc_feature    complement(30032. .30494)
                     /note="match: GSS: Em:AQ723457"
     repeat_region   31501. .31807
                     /note="AluSx repeat: matches 1. .307 of consensus"
     repeat_region   32165. .32267
                     /note="L2 repeat: matches 2554. .2646 of consensus"
     repeat_region   32268. .32752
                     /note="MER66B repeat: matches 1. .486 of consensus"
     repeat_region   32753. .33089
                     /note="L2 repeat: matches 2213. .2554 of consensus"
     repeat_region   33351. .33651
                     /note="AluSx repeat: matches 1. .304 of consensus"
     repeat_region   33662. .33966
                     /note="AluY repeat: matches 1. .295 of consensus"
     repeat_region   34730. .34818
                     /note="MIR repeat: matches 43. .145 of consensus"
     misc_feature    complement(34862. .35287)
                     /note="match: GSS: Em:AQ404224"
     misc_feature    complement(35239. .35778)
                     /note="match: GSS: Em:AQ634531"
     misc_feature    complement(35374. .35798)
                     /note="match: GSS: Em:AQ565906"
     misc_feature    36173. .36533
                     /note="match: GSS: Em:AQ231637"
     repeat_region   36392. .36553
                     /note="MER5A repeat: matches 4. .175 of consensus"
     repeat_region   36668. .36961
                     /note="AluSg repeat: matches 1. .297 of consensus"
     repeat_region   37033. .37258
                     /note="MIR repeat: matches 15. .252 of consensus"
     misc_feature    38080. .38507
                     /note="match: GSS: Em:B36377"
     misc_feature    38087. .38612
                     /note="match: GSS: Em:AQ680468"
```

```
     repeat_region   38349. .38418
                     /note="35 copies 2 mer tg 74% conserved"
     repeat_region   38694. .38979
                     /note="AluJo repeat: matches 1. .286 of consensus"
     misc_feature    40625. .41129
                     /note="match: GSS: Em:AQ680145"
     repeat_region   41704. .41819
                     /note="MIR repeat: matches 57. .177 of consensus"
     repeat_region   41842. .42034
                     /note="L1MC4 repeat: matches 7779. .7969 of consensus"
     repeat_region   42037. .42226
                     /note="MLT1J repeat: matches 45. .247 of consensus"
     misc_feature    complement(42848. .43207)
                     /note="match: GSS: Em:AQ385536"
     repeat_region   42966. .43258
                     /note="L2 repeat: matches 1138. .1493 of consensus"
     repeat_region   43259. .43555
                     /note="AluJb repeat: matches 10. .303 of consensus"
     repeat_region   43556. .43741
                     /note="L2 repeat: matches 970. .1138 of consensus"
BASE COUNT    12960 a   8117 c   8797 g  13867 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 9;  Length 43741;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattatt 24
           ||||||||||||||||
Db   10623 AAGTCACAGAATTATT 10608


RESULT  34
AC074073/c
LOCUS       AC074073    68894 bp    DNA             HTG       13-JUL-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-791I6 map 3, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC074073
VERSION     AC074073.1  GI:9170900
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68894)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-791I6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 68894)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7127
                Center clone name: 791_I_6
            --------------.
            * NOTE: This record contains 83 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      747: contig of 747 bp in length
            *      748 847: gap of      100 bp
            *      848     1579: contig of 732 bp in length
            *     1580 1679: gap of      100 bp
            *     1680     2425: contig of 746 bp in length
            *     2426 2525: gap of      100 bp
            *     2526     3254: contig of 729 bp in length
            *     3255 3354: gap of      100 bp
            *     3355     4075: contig of 721 bp in length
            *     4076 4175: gap of      100 bp
            *     4176     4892: contig of 717 bp in length
            *     4893 4992: gap of      100 bp
            *     4993     5717: contig of 725 bp in length
            *     5718 5817: gap of      100 bp
            *     5818     6558: contig of 741 bp in length
            *     6559 6658: gap of      100 bp
            *     6659     7371: contig of 713 bp in length
            *     7372 7471: gap of      100 bp
            *     7472     8209: contig of 738 bp in length
            *     8210 8309: gap of      100 bp
            *     8310     9046: contig of 737 bp in length
            *     9047 9146: gap of      100 bp
            *     9147     9888: contig of 742 bp in length
            *     9889 9988: gap of      100 bp
            *     9989    10723: contig of 735 bp in length
            *    10724 10823: gap of      100 bp
            *    10824    11553: contig of 730 bp in length
            *    11554 11653: gap of      100 bp
            *    11654    12391: contig of 738 bp in length
            *    12392 12491: gap of      100 bp
            *    12492    13214: contig of 723 bp in length
            *    13215 13314: gap of      100 bp
            *    13315    14034: contig of 720 bp in length
            *    14035 14134: gap of      100 bp
            *    14135    14864: contig of 730 bp in length
            *    14865 14964: gap of      100 bp
            *    14965    15700: contig of 736 bp in length
            *    15701 15800: gap of      100 bp
            *    15801    16535: contig of 735 bp in length
            *    16536 16635: gap of      100 bp
            *    16636    17357: contig of 722 bp in length
            *    17358 17457: gap of      100 bp
            *    17458    18192: contig of 735 bp in length
            *    18193 18292: gap of      100 bp
            *    18293    19030: contig of 738 bp in length
            *    19031 19130: gap of      100 bp
            *    19131    19857: contig of 727 bp in length
            *    19858 19957: gap of      100 bp
            *    19958    20677: contig of 720 bp in length
```

```
*    20678 20777: gap of      100 bp
*    20778    21514: contig of 737 bp in length
*    21515 21614: gap of      100 bp
*    21615    22336: contig of 722 bp in length
*    22337 22436: gap of      100 bp
*    22437    23190: contig of 754 bp in length
*    23191 23290: gap of      100 bp
*    23291    24024: contig of 734 bp in length
*    24025 24124: gap of      100 bp
*    24125    24859: contig of 735 bp in length
*    24860 24959: gap of      100 bp
*    24960    25700: contig of 741 bp in length
*    25701 25800: gap of      100 bp
*    25801    26544: contig of 744 bp in length
*    26545 26644: gap of      100 bp
*    26645    27367: contig of 723 bp in length
*    27368 27467: gap of      100 bp
*    27468    28196: contig of 729 bp in length
*    28197 28296: gap of      100 bp
*    28297    29028: contig of 732 bp in length
*    29029 29128: gap of      100 bp
*    29129    29843: contig of 715 bp in length
*    29844 29943: gap of      100 bp
*    29944    30633: contig of 690 bp in length
*    30634 30733: gap of      100 bp
*    30734    31472: contig of 739 bp in length
*    31473 31572: gap of      100 bp
*    31573    32319: contig of 747 bp in length
*    32320 32419: gap of      100 bp
*    32420    33151: contig of 732 bp in length
*    33152 33251: gap of      100 bp
*    33252    33977: contig of 726 bp in length
*    33978 34077: gap of      100 bp
*    34078    34818: contig of 741 bp in length
*    34819 34918: gap of      100 bp
*    34919    35632: contig of 714 bp in length
*    35633 35732: gap of      100 bp
*    35733    36474: contig of 742 bp in length
*    36475 36574: gap of      100 bp
*    36575    37300: contig of 726 bp in length
*    37301 37400: gap of      100 bp
*    37401    38137: contig of 737 bp in length
*    38138 38237: gap of      100 bp
*    38238    38960: contig of 723 bp in length
*    38961 39060: gap of      100 bp
*    39061    39789: contig of 729 bp in length
*    39790 39889: gap of      100 bp
*    39890    40617: contig of 728 bp in length
*    40618 40717: gap of      100 bp
*    40718    41442: contig of 725 bp in length
*    41443 41542: gap of      100 bp
*    41543    42278: contig of 736 bp in length
*    42279 42378: gap of      100 bp
*    42379    43117: contig of 739 bp in length
*    43118 43217: gap of      100 bp
*    43218    43961: contig of 744 bp in length
*    43962 44061: gap of      100 bp
*    44062    44786: contig of 725 bp in length
*    44787 44886: gap of      100 bp
*    44887    45614: contig of 728 bp in length
*    45615 45714: gap of      100 bp
*    45715    46458: contig of 744 bp in length
*    46459 46558: gap of      100 bp
*    46559    47292: contig of 734 bp in length
*    47293 47392: gap of      100 bp
*    47393    48107: contig of 715 bp in length
*    48108 48207: gap of      100 bp
*    48208    48935: contig of 728 bp in length
*    48936 49035: gap of      100 bp
*    49036    49765: contig of 730 bp in length
*    49766 49865: gap of      100 bp
*    49866    50604: contig of 739 bp in length
*    50605 50704: gap of      100 bp
*    50705    51434: contig of 730 bp in length
*    51435 51534: gap of      100 bp
*    51535    52266: contig of 732 bp in length
*    52267 52366: gap of      100 bp
*    52367    53096: contig of 730 bp in length
*    53097 53196: gap of      100 bp
*    53197    53936: contig of 740 bp in length
*    53937 54036: gap of      100 bp
*    54037    54770: contig of 734 bp in length
*    54771 54870: gap of      100 bp
*    54871    55574: contig of 704 bp in length
*    55575 55674: gap of      100 bp
*    55675    56392: contig of 718 bp in length
*    56393 56492: gap of      100 bp
*    56493    57227: contig of 735 bp in length
*    57228 57327: gap of      100 bp

  Query Match             64.0%;  Score 16;  DB 2;  Length 68894;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db   58360 AGTCACAGAATTATTT 58345


RESULT  35
AL590387/c
LOCUS       AL590387    76075 bp    DNA             PRI       06-JUN-2001
DEFINITION  Human DNA sequence from clone RP3-356B20 on chromosome 6, complete
            sequence.
ACCESSION   AL590387
VERSION     AL590387.7  GI:14330109
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 76075)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 8, 2001 this sequence version replaced gi:14141468.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP3-356B20 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            IMPORTANT: This sequence is not the entire insert of clone
            RP3-356B20 It may be shorter because we sequence overlapping
```

```
          sections only once, except for a 100 base overlap.
          The true left end of clone RP11-505E17 is at 75976 in this
          sequence. The true right end of clone RP5-1043E3 is at 100 in this
          sequence.
FEATURES             Location/Qualifiers
     source          1. .76075
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP3-356B20"
                     /clone_lib="RPCI-3"
     repeat_region   1297. .1521
                     /note="MIR repeat: matches 1. .252 of consensus"
     repeat_region   1916. .2364
                     /note="L2 repeat: matches 616. .1112 of consensus"
     repeat_region   2364. .2716
                     /note="L2 repeat: matches 1303. .1623 of consensus"
     repeat_region   2717. .3010
                     /note="AluSx repeat: matches 1. .297 of consensus"
     repeat_region   3011. .3967
                     /note="L2 repeat: matches 1623. .2711 of consensus"
     repeat_region   4896. .4947
                     /note="13 copies 4 mer atat 82% conserved"
     repeat_region   4907. .4948
                     /note="21 copies 2 mer ta 78% conserved"
     repeat_region   4949. .5247
                     /note="AluSq repeat: matches 1. .299 of consensus"
     repeat_region   5894. .5992
                     /note="MIR repeat: matches 114. .214 of consensus"
     repeat_region   6328. .6635
                     /note="AluY repeat: matches 1. .308 of consensus"
     repeat_region   7551. .7704
                     /note="L2 repeat: matches 2586. .2750 of consensus"
     repeat_region   8217. .8565
                     /note="L2 repeat: matches 2381. .2708 of consensus"
     repeat_region   9314. .9355
                     /note="21 copies 2 mer tt 76% conserved"
     repeat_region   9799. .9969
                     /note="MIR repeat: matches 13. .191 of consensus"
     repeat_region   10000. .10467
                     /note="L1MB4 repeat: matches 5692. .6183 of consensus"
     repeat_region   11167. .11722
                     /note="L1M4 repeat: matches 4081. .4646 of consensus"
     repeat_region   11822. .12380
                     /note="L1M4 repeat: matches 4681. .5285 of consensus"
     repeat_region   12381. .12700
                     /note="AluSx repeat: matches 1. .299 of consensus"
     repeat_region   12701. .12821
                     /note="L1M4 repeat: matches 5285. .5398 of consensus"
     repeat_region   13075. .13130
                     /note="28 copies 2 mer tg 76% conserved"
     repeat_region   14836. .14924
                     /note="L2 repeat: matches 2583. .2675 of consensus"
     repeat_region   14993. .15199
                     /note="L2 repeat: matches 1365. .1575 of consensus"
     repeat_region   15295. .15355
                     /note="MIR repeat: matches 20. .80 of consensus"
     repeat_region   15741. .15915
                     /note="L2 repeat: matches 1994. .2171 of consensus"
     repeat_region   16209. .16244
                     /note="9 copies 4 mer acac 97% conserved"
     repeat_region   16938. .17014
                     /note="L2 repeat: matches 2575. .2662 of consensus"
     misc_feature    17193. .17608
                     /note="match: GSS: Em:AQ023146"
     repeat_region   20366. .20409
                     /note="2 copies 22 mer 97% conserved"
     repeat_region   20561. .20778
                     /note="109 copies 2 mer tt 56% conserved"
     repeat_region   20622. .20669
                     /note="16 copies 3 mer cct 77% conserved"
     misc_feature    21469. .21519
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
     repeat_region   22092. .22221
                     /note="FLAM_A repeat: matches 1. .130 of consensus"
     repeat_region   22713. .23053
                     /note="MER47A repeat: matches 27. .366 of consensus"
     misc_feature    23361. .23811
                     /note="match: GSS: Em:B82113"
     repeat_region   23448. .23650
                     /note="MIR repeat: matches 12. .227 of consensus"
     repeat_region   25522. .25829
                     /note="AluSx repeat: matches 1. .306 of consensus"
     misc_feature    26053. .26338
                     /note="match: GSS: Em:B36212"
     repeat_region   26876. .26902
                     /note="9 copies 3 mer cac 92% conserved"
     repeat_region   27024. .27123
                     /note="MER53 repeat: matches 17. .122 of consensus"
     repeat_region   27146. .27222
                     /note="MIR repeat: matches 187. .262 of consensus"
     repeat_region   27435. .27548
                     /note="MIR repeat: matches 20. .137 of consensus"
     repeat_region   29525. .29681
                     /note="FRAM repeat: matches 1. .158 of consensus"
     misc_feature    29555. .30038
                     /note="match: GSS: Em:AQ589839"
     misc_feature    29556. .29942
                     /note="match: GSS: Em:AQ110308"
     misc_feature    29568. .29957
                     /note="match: GSS: Em:AQ193558"
     repeat_region   30020. .30140
                     /note="MER5B repeat: matches 1. .123 of consensus"
     misc_feature    30714. .31046
                     /note="match: GSS: Em:AZ515855"
     repeat_region   32032. .32147
                     /note="MER91B repeat: matches 14. .122 of consensus"
     repeat_region   32844. .33232
                     /note="MER49 repeat: matches 2. .405 of consensus"
     repeat_region   33264. .33467
                     /note="MER4D repeat: matches 689. .895 of consensus"
     misc_feature    33286. .33710
                     /note="match: GSS: Em:B66571"
     repeat_region   34843. .34960
                     /note="U1 repeat: matches 2. .135 of consensus"
     repeat_region   35256. .35565
                     /note="AluSx repeat: matches 1. .310 of consensus"
     repeat_region   35694. .35737
                     /note="11 copies 4 mer acac 93% conserved"
     repeat_region   35978. .36140
                     /note="L2 repeat: matches 2570. .2750 of consensus"
     repeat_region   36381. .36689
                     /note="AluSq repeat: matches 1. .308 of consensus"
     misc_feature    complement(37459. .37567)
                     /note="match: GSS: Em:AQ236853"
     misc_feature    37595. .38091
                     /note="match: GSS: Em:AQ127624"
     repeat_region   39369. .39730
                     /note="MLT1A2 repeat: matches 4. .374 of consensus"
     misc_feature    39839. .40080
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
     repeat_region   40326. .40593
                     /note="L2 repeat: matches 670. .941 of consensus"
     misc_feature    40351. .40368
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
     misc_feature    40815. .41335
                     /note="match: GSS: Em:AQ562847"
     repeat_region   40968. .41155
                     /note="MIR repeat: matches 51. .256 of consensus"
     repeat_region   41233. .41290
                     /note="MER5A repeat: matches 122. .179 of consensus"
     repeat_region   41283. .41365
                     /note="MER5B repeat: matches 96. .178 of consensus"
```

```
     repeat_region   41564. .42004
                     /note="L2 repeat: matches 2256. .2750 of consensus"
     repeat_region   42012. .42110
                     /note="L2 repeat: matches 2587. .2693 of consensus"
     repeat_region   42443. .42526
                     /note="42 copies 2 mer at 63% conserved"
     repeat_region   42537. .42787
                     /note="L1MD2 repeat: matches 6108. .6341 of consensus"
     repeat_region   42853. .43166

  Query Match             64.0%;  Score 16;  DB 9;  Length 76075;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db  17122 AGTCACAGAATTATTT 17107


RESULT  36
AC012679/c
LOCUS       AC012679   105937 bp    DNA             PLN       19-JAN-2001
DEFINITION  Arabidopsis thaliana chromosome 1 BAC F25P22 genomic sequence,
            complete sequence.
ACCESSION   AC012679
VERSION     AC012679.5  GI:12324196
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 105937)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC F25P22 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105937)
  AUTHORS   Lin,X. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE   3  (bases 1 to 105937)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Jan 19, 2001 this sequence version replaced gi:12280804.
            Address all correspondence to:at@tigr.org

            BAC clone F25P22 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES             Location/Qualifiers
     source          1. .105937
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="I"
                     /map="nga111"
                     /clone="F25P22"
     mRNA            join(<1. .76,158. .299,430. .710,789. .950,1031. .1084,
                     1171. .1512)
                     /gene="F25P22.1"
     gene            1. .1512
                     /gene="F25P22.1"
                     /note="similar to
                     S-adenosyl-methionine-sterol-C-methyltransferase
                     GB:AAC34951 [Nicotiana tabacum]"
     CDS             join(<1. .76,158. .299,430. .710,789. .950,1031. .1084,
                     1171. .1344)
                     /gene="F25P22.1"
                     /codon_start=2
                     /product="putative
                     S-adenosyl-methionine-sterol-C-methyltransferase, 5'
                     partial; 1-1344"
                     /protein_id="AAG52075.1"
                     /db_xref="GI:12324209"
                     /translation="NSYELSLVSCKCIGAYVRNKKNQNQICWLWQKVSSDNDRGFQRF
                     LDNVQYKSSGILRYERVFGEGFVSTGGLETTKEFVDMLDLKPGQKVLDVGCGIGGGDF
                     YMAENFDVDVVGIDLSVNMISFALEHAIGLKCSVEFEVADCTKKEYPDNTFDVIYSRD
                     TILHIQDKPALFRRFYKWLKPGGKVLITDYCRSPKTPSPDFAIYIKKRGYDLHDVQAY
                     GQMLRDAGFEEVIAEDRTDQFMKVLKRELDAVEKEKEEFISDFSKEDYEDIIGGWKSK
                     LLRSSSGEQKWGLFIAKRN"
     mRNA            join(<6321. .6579,6690. .6820,6910. .7146,7259. .7502,
                     7588. .>7751)
                     /gene="F25P22.2"
     gene            6321. .7751
                     /gene="F25P22.2"
                     /note="contains Pfam profile: PF00657 lipase/acylhydrolase
                     with GDSL-like motif"
     CDS             join(6321. .6579,6690. .6820,6910. .7146,7259. .7502,
                     7588. .7751)
                     /gene="F25P22.2"
                     /codon_start=1
                     /product="putative lipase/acylhydrolase; 6321-7751"
                     /protein_id="AAG52082.1"
                     /db_xref="GI:12324216"
                     /translation="MNCLMFFKMLLAFSFISLFYVGNAQQSYGNSTVSALFAFGDSIL
                     DTGNNNLLLSVSKVNFYPYGRDFIGGRATGRFGNGRVFSDIIAEGLGLKNLLPAYRDP
                     YLWNNDLTTGVCFASGGSGLDPITARTTGSIWVSDQVTDFQNYITRLNGVVGNQEQAN
                     AVISNAVYLISAGNNDIAITYFTTGARRLQYTLPAYNDQLVSWTRDLIKSLYDMGARK
                     FAVMGTLPLGCLPGARALTRACELFVNQGAAMFNQQLSADIDNLGATFPGAKFVYVDM
                     YNPLLGLIINPQASGFIDVADACCCTPTHLIPCLDASRYVFWDVAHPTQKSYETIAPQ
                     IIENIKAKLA"
     mRNA            join(<9376. .9496,9643. .9967,10550. .>10898)
                     /gene="F25P22.3"
     gene            9376. .10898
                     /gene="F25P22.3"
                     /note="contains Pfam profile: PF00314 thaumatin family"
     CDS             join(9376. .9496,9643. .9967,10550. .10898)
                     /gene="F25P22.3"
                     /codon_start=1
                     /product="thaumatin-like protein; 9376-10898"
                     /protein_id="AAG52086.1"
                     /db_xref="GI:12324220"
                     /translation="MTNEKCEVALPLKLKSEKPSMASSSSLFLLPLLLLVSQASASTV
                     IFYNKCTYTVWPGIQPSSGQSLLAGGGFKLSPNRAYTLQLPPLWSGRFWGRHGCSFDR
                     SGRGRCATGDCGGSFLCNGAGGVPPATLAEITLGHDMDFYDVSLVDGYNLAMSIMPVK
                     GTGKCTYAGCVSDLNRMCPVGLQVRSRDGTQVVACKSACSAFNSPRYCCTGLFGNPQS
                     CKPTAYSKIFKVACPKAYSYAYDDPTSIATCSKANYVVTFCPHRGR"
     mRNA            <12692. .>13183
                     /gene="F25P22.4"
     gene            12692. .13183
                     /gene="F25P22.4"
```

```
                     /note="similar to calmodulin GB:P02596 [Renilla
                     reniformis]; contains Pfam profile: PF00036 EF hand (4
                     copies)"
     CDS             12692. .13183
                     /gene="F25P22.4"
                     /codon_start=1
                     /product="putative calmodulin; 12692-13183"
                     /protein_id="AAG52088.1"
                     /db_xref="GI:12324222"
                     /translation="MANTNLESTNKSTTPSTDMELKKVFDKFDANGDGKISVSELGNV
                     FKSMGTSYTEEELNRVLDEIDIDCDGFINQEEFATICRSSSSAVEIREAFDLYDQNKN
                     GLISSSEIHKVLNRLGMTCSVEDCVRMIGHVDTDGDGNVNFEEFQKMMSSPELVKGTV
                     ANS"
     mRNA            join(<14977. .15197,15451. .>15931)
                     /gene="F25P22.5"
     gene            14977. .15931
                     /gene="F25P22.5"
                     /note="contains Pfam profile: PF00071 ras family"
     CDS             join(14977. .15197,15451. .15931)
                     /gene="F25P22.5"
                     /codon_start=1
                     /product="putative ras-related GTP-binding protein;
                     14977-15931"
                     /protein_id="AAG52089.1"
                     /db_xref="GI:12324223"
                     /translation="MAEDTYEEECDYLFKAVLIGDSAVGKSNLLSRFSKDEFRFDSKP
                     TIGVEFAYRNVHVGDKIIKAQIWDTAGQERFRAITSSYYRGALGALLIYDITRRTTFD
                     NIKKWLFELRDFANPETVVVLVGNKSDLRQSREVEEDEGKTLAESEGLYFLETSALEN
                     VNVEEAFLVMIGRIHEVVTQRIASENKSNGAATPHINGNGNGTVLPVGKEIVNIHEVT
                     ATQPLLSSSSNCCFK"
     mRNA            complement(join(<16797. .16967,17064. .17297,17388. .17469,
                     17546. .17607,17706. .17773,17866. .17935,18448. .18772,
                     19179. .19223,19427. .>19725))
                     /gene="F25P22.7"
     gene            complement(16797. .19725)
                     /gene="F25P22.7"
     CDS             complement(join(16797. .16967,17064. .17297,17388. .17469,
                     17546. .17607,17706. .17773,17866. .17935,18448. .18772,
                     19179. .19223,19427. .19725))
                     /gene="F25P22.7"
                     /codon_start=1
                     /product="unknown protein; 19725-16797"
                     /protein_id="AAG52066.1"
                     /db_xref="GI:12324200"
                     /translation="MATLFTATVPSHHRFVSPSQHPKQSLLSQSLSVTFTENPQPTAV
                     VTLQEQQLTDWITSPVTRRFGIGAGFTWAGFLAFGVVSEQMKKSRLDVFQEEDNTRGL
                     EKQEEIILPNGIRYYDLQVGSGATPSSGYLVVFDVKGQVHGTEQVFVDTFGGKGKSLA
                     MVMDSRPYSKGLCQGIEHVLRSMKAGGKRKVIIPPSLGFGDRNVEFGQGLEIPPSATL
                     DYIIEVIYQFIFFVITALFKFDQVTDFAGSTNFVILAVLTLVLKASWHFRQIVLTLLV
                     VVWGLRLGIFLLMRILQWGEDRRFDEQRGNIVRLIIFWTLQAVWVWTVSLPLTLVNAS
                     DGGGSLKPADVIGWTMWVFGFLIEAAADQQKLSFKNSPENRGKWCDVGVWKYSRHPNY
                     FGEMLLWWGIFVAASPVLEGAEYLVIFGPLFLTLLLLFVSGIPLLEVLPCSLPLYLDN
                     F"
     mRNA            complement(join(<20191. .20329,20451. .20627,20721. .20770,
                     20870. .20938,21048. .21146,21239. .21317,21582. .21682,
                     21776. .21844,21957. .21997,22081. .23243,23382. .23485,
                     23568. .24008,24102. .>24662))
                     /gene="F25P22.8"
     gene            complement(20191. .24662)
                     /gene="F25P22.8"
                     /note="contains Pfam profile: PF00069 eukaryotic protein
                     kinase domain"
     CDS             complement(join(20191. .20329,20451. .20627,20721. .20770,
                     20870. .20938,21048. .21146,21239. .21317,21582. .21682,
                     21776. .21844,21957. .21997,22081. .23243,23382. .23485,
                     23568. .24008,24102. .24662))
                     /gene="F25P22.8"
                     /codon_start=1
                     /product="putative protein kinase; 24662-20191"
                     /protein_id="AAG52069.1"
                     /db_xref="GI:12324203"
                     /translation="MKVKEETLKNLGDGVVLRPVDHCSSIWSMKMNMKNFLKKLHISP
                     NQSDEAEGSISTTKSNHHKSIDVSSSSSPRSHHSNSPEIKPFSGLSNWLSSVGHRKIP
```

```
                     SPPNSFNAKNRAATVDDTVVVNGSEHVDLGSKDPAVEEENQIQLALELSAREDPEATQ
                     IEAIKQFSLGSCAPENSPAELIAYRYWNYNCLGYDDKILDGFYDLYGVLNASSAERIP

       Query Match             64.0%;  Score 16;  DB 8;  Length 105937;
       Best Local Similarity   100.0%;  Pred. No. 13;
       Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

     Qy      10 agtcacagaattattt 25
                ||||||||||||||||
     Db   55362 AGTCACAGAATTATTT 55347


     RESULT  37
     AC044784_1/c
     WPCOMMENT
       Sequence split into 4 fragments  LOCUS AC044784 Accession AC044784
          Fragment Name      Begin     End
          AC044784_0             1     110000
          AC044784_1        100001     210000
          AC044784_2        200001     310000
          AC044784_3        300001     365176
     Continuation (2 of 4) of AC044784 from base 100001 (AC044784 Homo sapiens chromosome


       Query Match             64.0%;  Score 16;  DB 2;  Length 110000;
       Best Local Similarity   100.0%;  Pred. No. 13;
       Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

     Qy       8 caagtcacagaattat 23
                ||||||||||||||||
     Db   10285 CAAGTCACAGAATTAT 10270


     RESULT  38
     AB020878
     LOCUS       AB020878   110525 bp    DNA             PRI       21-MAY-1999
     DEFINITION  Homo sapiens genomic DNA of 9q32 anti-oncogene of flat epitherium
                 cancer , segment 10/10.
     ACCESSION   AB020878
     VERSION     AB020878.1  GI:4003398
     KEYWORDS    .
     SOURCE      Homo sapiens DNA.
       ORGANISM  Homo sapiens
                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
     REFERENCE   1  (sites)
       AUTHORS   Nakamura,Y., Isomura,M. and Daigo,Y.
       TITLE     Homo sapiens 1,010,525bp genomic DNA of 9q32 anti-oncogene of flat
                 epitherium cancer region
       JOURNAL   Published Only in DataBase (1998) In press
     REFERENCE   2  (bases 1 to 110525)
       AUTHORS   Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
       TITLE     Direct Submission
       JOURNAL   Submitted (30-NOV-1998) to the DDBJ/EMBL/GenBank databases. Mika
                 Hirakawa, Japan Science and Technology Corporation (JST), Advanced
                 Database Department; 5-3, Yonban-cho, Chiyoda-ku, Tokyo 102-0028,
                 Japan (E-mail:mika@tokyo.jst.go.jp, Tel:81-3-5214-8491,
                 Fax:81-3-5214-8470)
     COMMENT     This sequence is conducted by Japanese Foundation for Cancer
                 Research as a JST seauencing team
                 Principal Investigator: Yusuke Nakamura Ph.D.
                 Phone:+81-3-5449-5372, Fax:+81-3-5449-5433,
                 yusuke@hgc.ims.u-tokyo.ac.jp
                 The sequence is submitted by:Human Genome Sequencing in ALIS
                 project of JST
                 Japan Science and Technology Corporation (JST)
                 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
                 For further infomation about this sequence, including its location
                 and relationship to other sequences, please visit our sequence
                 archive Web site
                 (http://www-alis.tokyo.jst.go.jp/HGS/top.html)
                 or send email to webmaster@www-alis.tokyo.jst.go.jp.
     FEATURES             Location/Qualifiers
```

```
     source          1. .110525
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="9q32"
     STS             80338. .80503
                     /standard_name="D9S1126"
                     /note="CHLC.GGAT11B01.P33342;The location is between each
                     flanking site of PCR primers."
                     /db_xref="GDB:686181"
     STS             86433. .86567
                     /note="stSG8871;The location is between each flanking site
                     of PCR primers."
                     /db_xref="GDB:4569679"
     STS             104545. .104691
                     /note="SHGC-18832;The location is between each flanking
                     site of PCR primers."
                     /db_xref="GDB:6458849"
BASE COUNT    32237 a  22817 c  23211 g  32215 t     45 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 9;  Length 110525;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattatt 24
           ||||||||||||||||
Db   39879 AAGTCACAGAATTATT 39894


RESULT  39
AC011248
LOCUS       AC011248   111312 bp    DNA             PRI       28-APR-2000
DEFINITION  Homo sapiens 12 BAC RP11-14P21 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC011248
VERSION     AC011248.8  GI:7658299
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 111312)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y.,
            Jones,M., Kelly,S., Kneitz,S., Kondejewski,N., Kong,Y., Kovar,C.,
            Lau,S., Leal,B., Lee,E., Li,Z., Lichtarge,O., Liu,J., Liu,W.,
            Logan,O., Lozado,R.J., Lu,J., Lucier,R., Marondel,I., Martin,R.,
            Martinez,C., McLeod,M.P., Mei,G., Merscher,S., Miller,A.,
            Montgomery,K.T., Morgan,M., Morris,S., Nash,S., Nelson,A.,
            Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S.,
            Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M.,
            Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Shim,C., Simon,M.,
            Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T., Vasquez,L.,
            Vinson,R., Vo,Q., Wahbah,M., Watlington,S., Weinstock,G.,
            Weinstock,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G.,
            Xhang,A.M., Yang,R., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 111312)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 111312)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 28, 2000 this sequence version replaced gi:6996901.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

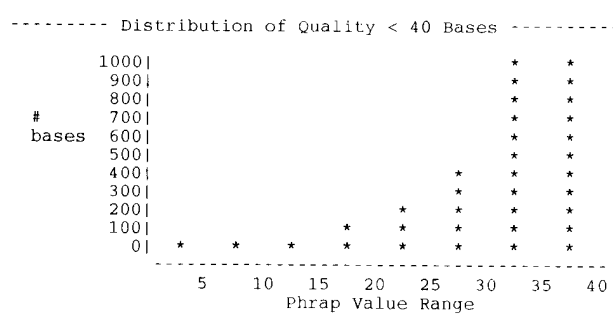
            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.
            html.

            QUALSTAT-REPORT----------


            ---------- Summary Statistics ------------------------------
            Contig length:                                  150286
            Phrap values in estimate:                       149476
            Average error rate (BCM-Phrap estimate):        0.000180006
            Fraction of Phrap values less than 40 :         0.0291418
            Number of consensus changing edits:             15
            Number of N's in consensus :                    1


            ---------- Consensus changing edits ------------------------
            Position      Original+Context          Edited+Context
            22441         aaagagaaga(n)gagggctcca   aaagagaaga(a)gagggctcca
            79231         agtaaggcca(n)annnttantn   agtaaggcca(t)atttttaatg
            79233         taaggccana(n)ntttantncc   taaggccata(t)ttttaatgcc
            79234         aaggccanan(n)tttantncct   aaggccatat(t)tttaatgcct
            79239         cananntta(n)tncctngtga    catattttta(a)tgcctcgtga
            79241         nannttant(n)cctngtgaaa    tatttttaat(g)cctcgtgaaa
            79245         tttantncct(n)gtgaaaataa   tttaatgcct(c)gtgaaaataa
            79279         agagctgtga(n)gctagaaatg   agagctgtga(a)gctagaaatg
            79440         agcctggcca(n)catcgtgaaa   agcctggcca(a)catcgtgaaa
            89920         tctcaaaaaa(n)caaaaagaag   tctcaaaaaa(a)caaaaagaag
            94117         gttttcccag(n)tcctggcaca   gttttcccag(t)tcctggcaca
            95790         gcgagactcc(n)tctcaaaaaa   gcgagactcc(a)tctcaaaaaa
            111017        ttaaggttcc(n)tcagctgaac   ttaaggttcc(a)tcagctgaac
            115887        atgccatata(n)caacaaacaa   atgccatata(a)caacaaacaa
            129846        ggaggctgaa(n)cgagtggatc   ggaggctgaa(g)cgagtggatc
```



Version: 1.01 qxfo.

```
FEATURES             Location/Qualifiers
     source          1. .111312
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-14P21"
     repeat_region   1. .133
                     /rpt_family="AluSg/x"
     repeat_region   167. .193
                     /rpt_family="(GAGAA)n"
     repeat_region   553. .795
                     /rpt_family="MIR"
     repeat_region   complement(824. .901)
                     /rpt_family="MIR"
     repeat_region   946. .985
                     /rpt_family="(CA)n"
     repeat_region   complement(1467. .1648)
                     /rpt_family="MLT1G"
     repeat_region   1876. .2176
                     /rpt_family="AluSq"
     repeat_region   complement(2180. .2431)
                     /rpt_family="AluSx"
     repeat_region   complement(2432. .2788)
                     /rpt_family="THE1B"
     repeat_region   complement(2789. .2835)
                     /rpt_family="AluSx"
     repeat_region   complement(2864. .3025)
                     /rpt_family="MLT1F1"
     repeat_region   3101. .3327
                     /rpt_family="MIR"
     repeat_region   complement(3383. .3455)
                     /rpt_family="L2"
     STS             3974. .4064
                     /standard_name="G07716"
                     /db_xref="dbSTS:13562"
     repeat_region   5179. .5480
                     /rpt_family="AluSp"
     repeat_region   complement(6278. .6347)
                     /rpt_family="MER4A"
     repeat_region   6348. .6648
                     /rpt_family="AluSq"
     repeat_region   complement(6649. .7184)
                     /rpt_family="MER4A"
     repeat_region   8486. .8703
                     /rpt_family="L2"
     repeat_region   8729. .9122
                     /rpt_family="L2"
     repeat_region   9213. .9497
                     /rpt_family="L2"
     repeat_region   9498. .9798
                     /rpt_family="AluY"
     repeat_region   9799. .9887
                     /rpt_family="L2"
     repeat_region   complement(10707. .12039)
                     /rpt_family="L1MA4"
     repeat_region   12047. .12155
                     /rpt_family="L2"
     repeat_region   complement(12165. .12729)
                     /rpt_family="L2"
     repeat_region   12702. .12847
                     /rpt_family="L2"
     STS             13332. .13486
                     /standard_name="Cda1fh02"
                     /db_xref="GDB:446262"
     repeat_region   13505. .13528
```

Query Match 64.0%; Score 16; DB 9; Length 111312;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       9 aagtcacagaattatt 24
           ||||||||||||||||
Db   86830 AAGTCACAGAATTATT 86845
```

```
RESULT  40
HSDJ655C5/c
LOCUS       HSDJ655C5  119733 bp    DNA             PRI       03-JUL-2001
DEFINITION  Human DNA sequence from clone RP4-655C5 on chromosome 6, complete
            sequence.
ACCESSION   AL121956
VERSION     AL121956.21  GI:14599397
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 119733)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 4, 2001 this sequence version replaced gi:14456158.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP4-655C5 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP4-655C5.
FEATURES             Location/Qualifiers
     source          1. .119733
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP4-655C5"
                     /clone_lib="RPCI-4"
repeat_region        1. .433
                     /note="L2 repeat: matches 1703. .2138 of consensus"
repeat_region        460. .950
                     /note="L2 repeat: matches 1042. .1597 of consensus"
misc_feature         complement(1180. .1481)
                     /note="match: STS: Em:HS168L15S"
repeat_region        1869. .2002
                     /note="L2 repeat: matches 2608. .2750 of consensus"
repeat_region        2158. .2498
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region        2774. .2978
                     /note="L2 repeat: matches 1337. .1539 of consensus"
repeat_region        3085. .3300
                     /note="MER76 repeat: matches 34. .257 of consensus"
repeat_region        3307. .3425
                     /note="MLT1A1 repeat: matches 1. .101 of consensus"
repeat_region        3444. .3662
                     /note="AluJb repeat: matches 1. .219 of consensus"
repeat_region        3666. .3842
                     /note="MLT1A1 repeat: matches 156. .341 of consensus"
misc_feature         complement(3839. .4322)
                     /note="match: GSS: Em:AZ419683"
misc_feature         complement(4102. .4359)
                     /note="match: GSS: Em:AZ639762"
misc_feature         4153. .4623
                     /note="match: GSS: Em:B87991"
misc_feature         4160. .4280
                     /note="match: GSS: Em:AZ264436"
misc_feature         4168. .4466
                     /note="match: GSS: Em:AQ008901"
misc_feature         complement(4217. .4491)
                     /note="match: GSS: Em:AQ207441"
misc_feature         complement(4225. .4551)
                     /note="match: GSS: Em:B46380"
misc_feature         4287. .4627
                     /note="match: GSS: Em:AZ264436"
misc_feature         4342. .4466
                     /note="match: GSS: Em:AZ266010"
misc_feature         complement(4430. .4627)
                     /note="match: GSS: Em:AZ324123"
misc_feature         4473. .4922
                     /note="match: GSS: Em:AZ266010"
repeat_region        4631. .4758
                     /note="MER85 repeat: matches 1. .139 of consensus"
misc_feature         complement(4759. .5131)
                     /note="match: GSS: Em:AQ773957"
misc_feature         complement(4759. .5141)
                     /note="match: GSS: Em:B80743"
misc_feature         4840. .5087
                     /note="match: GSS: Em:B81269"
misc_feature         4840. .5235
                     /note="match: GSS: Em:AQ226387"
misc_feature         4995. .5171
                     /note="match: GSS: Em:AZ361655"
misc_feature         complement(5054. .5532)
                     /note="match: GSS: Em:AQ636693"
misc_feature         5199. .5586
                     /note="match: GSS: Em:AZ361655"
misc_feature         complement(5300. .5752)
                     /note="match: GSS: Em:AZ246810"
repeat_region        5749. .6000
                     /note="AluJo repeat: matches 1. .254 of consensus"
repeat_region        6027. .6060
                     /note="MER76 repeat: matches 1. .34 of consensus"
repeat_region        6061. .6105
                     /note="L2 repeat: matches 2645. .2700 of consensus"
repeat_region        6113. .6160
                     /note="4 copies 12 mer 83% conserved"
repeat_region        6115. .6154
                     /note="10 copies 4 mer tgga 87% conserved"
repeat_region        7753. .8033
                     /note="AluSx repeat: matches 1. .289 of consensus"
repeat_region        8499. .8601
                     /note="L1MA8 repeat: matches 6184. .6286 of consensus"
repeat_region        9640. .9908
                     /note="L1MC5 repeat: matches 7609. .7898 of consensus"
misc_feature         complement(10367. .10902)
                     /note="match: GSS: Em:AZ520653"
repeat_region        11002. .11211
                     /note="MIR repeat: matches 31. .259 of consensus"
repeat_region        11273. .11378
                     /note="L2 repeat: matches 2543. .2658 of consensus"
repeat_region        11465. .11648
                     /note="MLT1A1 repeat: matches 222. .394 of consensus"
repeat_region        12112. .12163
                     /note="26 copies 2 mer tt 71% conserved"
repeat_region        12216. .12511
                     /note="AluSq repeat: matches 1. .299 of consensus"
repeat_region        13635. .13818
                     /note="MIR repeat: matches 58. .241 of consensus"
repeat_region        14327. .14590
                     /note="3 copies 88 mer 92% conserved"
repeat_region        16291. .16453
                     /note="AluSg/x repeat: matches 135. .297 of consensus"
repeat_region        16454. .16489
                     /note="Tigger4(Zombi) repeat: matches 45. .80 of consensus"
repeat_region        16518. .16607
                     /note="Tigger4(Zombi) repeat: matches 2642. .2731 of
                     consensus"
misc_feature         complement(16856. .17240)
                     /note="match: GSS: Em:AQ245004"
repeat_region        17145. .17319
                     /note="MER5A repeat: matches 1. .181 of consensus"
repeat_region        17602. .17902
                     /note="AluSc repeat: matches 2. .301 of consensus"
repeat_region        18131. .18420
                     /note="AluSx repeat: matches 1. .291 of consensus"
repeat_region        19188. .19469
                     /note="AluSx repeat: matches 1. .297 of consensus"
repeat_region        19497. .19762
                     /note="AluJo repeat: matches 1. .302 of consensus"
repeat_region        21340. .21430
                     /note="L2 repeat: matches 2564. .2667 of consensus"
repeat_region        21479. .21544
                     /note="2 copies 33 mer 93% conserved"
repeat_region        22660. .23022
                     /note="LTR14B repeat: matches 211. .608 of consensus"
repeat_region        23011. .23058
                     /note="LTR14B repeat: matches 1. .46 of consensus"
repeat_region        23126. .23881
                     /note="HERVK4 repeat: matches 5030. .5801 of consensus"
misc_feature         complement(23734. .23982)
                     /note="match: GSS: Em:AQ462686"
repeat_region        23931. .25376
                     /note="HERV17 repeat: matches 2365. .5466 of consensus"
repeat_region        25361. .26415
                     /note="HERV17 repeat: matches 985. .1987 of consensus"
repeat_region        26416. .26572
                     /note="Charlie1 repeat: matches 1. .177 of consensus"
repeat_region        26622. .26828
                     /note="Charlie1 repeat: matches 2506. .2757 of consensus"
repeat_region        26850. .27019
                     /note="MIR repeat: matches 23. .209 of consensus"
repeat_region        28015. .28168
                     /note="Tigger2a repeat: matches 273. .429 of consensus"
repeat_region        28169. .28475
                     /note="AluY repeat: matches 1. .311 of consensus"
misc_feature         complement(28448. .28942)
                     /note="match: GSS: Em:AQ663299"
repeat_region        28476. .28760
                     /note="Tigger2a repeat: matches 1. .273 of consensus"
repeat_region        29037. .29340
```

```
                     /note="AluSx repeat: matches 1. .303 of consensus"
     misc_feature    31526. .33561
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   36658. .36962
                     /note="AluY repeat: matches 1. .304 of consensus"
     repeat_region   37387. .37481

  Query Match             64.0%;  Score 16;  DB 9;  Length 119733;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaattatt 24
          ||||||||||||||||
Db   3994 AAGTCACAGAATTATT 3979


RESULT  41
AC008951/c
LOCUS       AC008951   124103 bp    DNA             HTG       01-JUL-2001
DEFINITION  Homo sapiens chromosome 5 clone CTD-2337A12, WORKING DRAFT
            SEQUENCE, 4 ordered pieces.
ACCESSION   AC008951
VERSION     AC008951.5  GI:14579694
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124103)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 124103)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 1, 2001 this sequence version replaced gi:7710864.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 756969
            Center clone name: CITB-H1_2337A12
            -------
            Summary Statistics
            Consensus quality: 120091 bases at least Q40
            Consensus quality: 122780 bases at least Q30
            Consensus quality: 123423 bases at least Q20
            Estimated insert size: 123000; pulse field gel estimation
            Estimated insert size: 123803; sum-of-contigs estimation
            Quality coverage:  5.58 in Q20 bases; pulse field gel estimation
            Quality coverage:  5.54 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    29001: contig of 29001 bp in length
            *    29002    29101: gap of unknown length
            *    29102    48451: contig of 19350 bp in length
            *    48452    48551: gap of unknown length
            *    48552    51607: contig of 3056 bp in length
            *    51608    51707: gap of unknown length
            *    51708   124103: contig of 72396 bp in length.
FEATURES             Location/Qualifiers
     source          1. .124103
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2337A12"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    34654 a  24183 c  24436 g  40530 t    300 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 124103;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 agtcacagaattattt 25
          ||||||||||||||||
Db  24194 AGTCACAGAATTATTT 24179


RESULT  42
AL159168
LOCUS       AL159168   129010 bp    DNA             PRI       30-SEP-2000
DEFINITION  Human DNA sequence from clone RP11-401H23 on chromosome 9, complete
            sequence.
ACCESSION   AL159168
VERSION     AL159168.15  GI:10443391
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129010)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 1, 2000 this sequence version replaced gi:10178449.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-401H23 is from the library RPCI-11.2 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-401H23 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-401H23 is at 1 in this sequence.
            The true left end of clone RP11-16L21 is at 128911 in this
            sequence. The true right end of clone RP11-386D8 is at 19333 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .129010
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-401H23"
                     /clone_lib="RPCI-11.2"
BASE COUNT    39045 a  27222 c  26654 g  36089 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 9;  Length 129010;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db 102456 AGTCACAGAATTATTT 102471


RESULT  43
AC019289/c
LOCUS       AC019289   139218 bp    DNA             HTG       08-APR-2000
DEFINITION  Homo sapiens chromosome 4 clone RP11-203I19 map 4, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC019289
VERSION     AC019289.3  GI:7528059
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139218)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 4, clone RP11-203I19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 139218)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 8, 2000 this sequence version replaced gi:6778536.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4306
                Center clone name: 203_I_19
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 132555 bases at least Q40
                Consensus quality: 135869 bases at least Q30
                Consensus quality: 137053 bases at least Q20
                Insert size: 140000; agarose-fp
                Insert size: 138118; sum-of-contigs
                Quality coverage: 4.9 in Q20 bases; agarose-fp
                Quality coverage: 5.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1      879: contig of 879 bp in length
            *     880 979: gap of      100 bp
            *     980     2319: contig of 1340 bp in length
            *    2320 2419: gap of      100 bp
            *    2420     6972: contig of 4553 bp in length
            *    6973 7072: gap of      100 bp
            *    7073    11233: contig of 4161 bp in length
            *   11234 11333: gap of      100 bp
            *   11334    18294: contig of 6961 bp in length
            *   18295 18394: gap of      100 bp
            *   18395    24919: contig of 6525 bp in length
            *   24920 25019: gap of      100 bp
            *   25020    32971: contig of 7952 bp in length
            *   32972 33071: gap of      100 bp
            *   33072    44845: contig of 11774 bp in length
            *   44846 44945: gap of      100 bp
            *   44946    60683: contig of 15738 bp in length
            *   60684 60783: gap of      100 bp
            *   60784    77274: contig of 16491 bp in length
            *   77275 77374: gap of      100 bp
            *   77375   102352: contig of 24978 bp in length
            *  102353 102452: gap of      100 bp
            *  102453   139218: contig of 36766 bp in length.
FEATURES             Location/Qualifiers
     source          1. .139218
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-203I19"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .879
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    980. .2319
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    2420. .6972
                     /note="assembly_fragment"
     misc_feature    7073. .11233
                     /note="assembly_fragment"
     misc_feature    11334. .18294
                     /note="assembly_fragment"
     misc_feature    18395. .24919
                     /note="assembly_fragment"
     misc_feature    25020. .32971
                     /note="assembly_fragment"
     misc_feature    33072. .44845
                     /note="assembly_fragment"
     misc_feature    44946. .60683
                     /note="assembly_fragment"
     misc_feature    60784. .77274
                     /note="assembly_fragment"
     misc_feature    77375. .102352
                     /note="assembly_fragment"
     misc_feature    102453. .139218
                     /note="assembly_fragment"
BASE COUNT    43319 a  25775 c  25817 g  43205 t   1102 others
```

```
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 139218;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           |||||||||||||||||
Db 103832 AGTCACAGAATTATTT 103817


RESULT  44
AC092893/c
LOCUS       AC092893   140002 bp    DNA             HTG       09-AUG-2001
DEFINITION  Homo sapiens chromosome 3q clone RP11-203I19, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC092893
VERSION     AC092893.1  GI:15134336
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 140002)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 140002)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     ---------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
            Center project name: HCSH
            Center clone name: RP11-203I19
            ---------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 134207 bases at least Q40
            Consensus quality: 137098 bases at least Q30
            Consensus quality: 138384 bases at least Q20
            Estimated insert size: 138999; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 4.8x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1    38256: contig of 38256 bp in length
            *     38257    38356: gap of unknown length
            *     38357    61565: contig of 23209 bp in length
            *     61566    61665: gap of unknown length
            *     61666    78299: contig of 16634 bp in length
            *     78300    78399: gap of unknown length
            *     78400    93197: contig of 14798 bp in length
            *     93198    93297: gap of unknown length
            *     93298   109082: contig of 15785 bp in length
            *    109083   109182: gap of unknown length
            *    109183   120959: contig of 11777 bp in length
            *    120960   121059: gap of unknown length
            *    121060   128609: contig of 7550 bp in length
            *    128610   128709: gap of unknown length
            *    128710   135281: contig of 6572 bp in length
            *    135282   135381: gap of unknown length
            *    135382   140002: contig of 4621 bp in length.
FEATURES             Location/Qualifiers
     source          1. .140002
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3q"
                     /clone="RP11-203I19"
BASE COUNT    44168 a  26034 c  26002 g  42981 t    817 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 140002;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           |||||||||||||||||
Db   1767 AGTCACAGAATTATTT 1752


RESULT  45
AC093258
LOCUS       AC093258   151206 bp    DNA             HTG       16-AUG-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-1E3, WORKING DRAFT SEQUENCE, 1
            unordered pieces.
ACCESSION   AC093258
VERSION     AC093258.1  GI:15193392
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151206)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151206)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 406848
            Center clone name: RPCI-11_1E3
            -------
            Summary Statistics
            Consensus quality: 150941 bases at least Q40
            Consensus quality: 151188 bases at least Q30
            Consensus quality: 151206 bases at least Q20
            Estimated insert size: 153880; agarose-fp estimation
            Estimated insert size: 151206; sum-of-contigs estimation
            Quality coverage:  10.02 in Q20 bases; agarose-fp estimation
            Quality coverage:  10.19 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   151206: contig of 151206 bp in length.
FEATURES              Location/Qualifiers
     source           1. .151206
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="5"
                      /clone="RP11-1E3"
                      /clone_lib="RPCI human BAC library 11"
BASE COUNT    49261 a  28930 c  27036 g  45979 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 151206;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db    9051 AGTCACAGAATTATTT 9066


Search completed: December 26, 2001, 12:50:42
Job time: 8764 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:16:28 ; Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196F-2
Perfect score: 25
Sequence: 1 attgtcgcaagtcacagaattattt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 39280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25	100.0	839	22	AAI62941	Human genomic DNA
2		25	100.0	1041	20	AAX01303	E. coli biotin syn
3		25	100.0	1084	10	AAN91329	E.coli Bio B gene.
4		25	100.0	1121	7	AAN60496	Sequence encoding
5		25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
c 6		17	68.0	17	14	AAQ56307	BioB DNA primer.
7		15	60.0	561	22	AAF94235	Primer specific fo
8		15	60.0	1771	20	AAZ30602	Mouse integrin alp
9		15	60.0	1771	22	AAS10264	Mouse integrin alp
c 10		15	60.0	3210	22	AAH18561	Human cDNA sequenc
c 11		15	60.0	3520	22	AAF93909	Human cDNA encodin
c 12		15	60.0	3520	22	AAF29361	Amyloid-beta prote
c 13		15	60.0	3748	22	AAF98729	Human late stage o
14		15	60.0	6087	22	AAS22683	Human cDNA encodin
15		15	60.0	6116	22	AAS22447	Human cDNA encodin
c 16		15	60.0	116624	19	AAV52850	Human eyal gene co
17		14	56.0	329	21	AAA41925	Human secreted exp
18		14	56.0	360	22	AAF65421	Novel human polynu
19		14	56.0	385	14	AAQ39649	Expressed Sequence
20		14	56.0	385	14	AAQ59061	Human brain Expres
21		14	56.0	403	22	AAI13743	Probe #3676 for ge
22		14	56.0	403	22	AAI35103	Probe #3789 used t
23		14	56.0	403	22	AAI03624	Probe #3615 used t
24		14	56.0	523	22	AAH12853	Human cDNA clone (
c 25		14	56.0	574	22	AAI41094	Probe #9780 used t
c 26		14	56.0	902	21	AAZ56756	Human transmembran
c 27		14	56.0	1050	9	AAN80927	Lividans Exported
c 28		14	56.0	1465	22	AAH14046	Human cDNA sequenc
c 29		14	56.0	1677	17	AAT31932	Retinoid X recepto
c 30		14	56.0	1787	17	AAT31928	Retinoid X recepto
c 31		14	56.0	1890	20	AAZ52940	Human prostate tum
32		14	56.0	1919	22	AAI60385	Human polynucleoti
c 33		14	56.0	2022	22	AAH15029	Human cDNA sequenc
c 34		14	56.0	2102	22	AAI58599	Human polynucleoti
c 35		14	56.0	2220	22	AAD12786	Human nuclear horm
c 36		14	56.0	2233	19	AAV10120	Human retinoid rec
c 37		14	56.0	2585	22	AAF77541	Caenorhabditis bri
c 38		14	56.0	2845	21	AAC69462	Human secreted pro
c 39		14	56.0	3413	20	AAV72330	F. balustinum CP70
c 40		14	56.0	5662	22	AAI61089	Human polynucleoti
41		14	56.0	5684	21	AAC59953	Human secreted pro
c 42		14	56.0	5773	21	AAC76014	Human ORFX ORF1569
c 43		14	56.0	5876	22	AAI59303	Human polynucleoti
c 44		14	56.0	7063	22	AAH24651	Nucleotide sequenc
45		14	56.0	8041	17	AAT10105	Adherence conferri
46		14	56.0	10126	22	AAI63610	Human kidney relat
47		14	56.0	14273	22	AAI62733	Human breast or ov
48		14	56.0	32351	21	AAF21307	Human low adenosin
49		14	56.0	32351	21	AAA35185	Human adenosine re
50		14	56.0	40298	21	AAF21311	Human low adenosin
51		14	56.0	40298	21	AAA35189	Human adenosine re
52		14	56.0	81001	22	AAF30035	Human apolipoprote
53		14	56.0	1038602	20	AAZ01425	Complete genome se
54		13	52.0	87	16	AAT22576	Human gene signatu
55		13	52.0	105	16	AAT20380	Human gene signatu
56		13	52.0	151	22	AAI26511	Probe #16444 for g
57		13	52.0	151	22	AAI55262	Probe #23948 used
c 58		13	52.0	167	16	AAT20233	Human gene signatu
c 59		13	52.0	187	21	AAC26213	Human secreted pro
60		13	52.0	259	22	AAI27363	Probe #17296 for g
61		13	52.0	259	22	AAI56259	Probe #24945 used
c 62		13	52.0	302	22	AAI23024	Probe #12957 for g
c 63		13	52.0	302	22	AAI48324	Probe #17010 used
c 64		13	52.0	302	22	AAI08683	Probe #8674 used t
65		13	52.0	325	22	AAI25914	Probe #15847 for g
66		13	52.0	325	22	AAI53184	Probe #21870 used
67		13	52.0	330	22	AAH35496	Human colon cancer
68		13	52.0	343	22	AAI63192	Human kidney relat
69		13	52.0	354	15	AAQ77082	Human genome fragm
c 70		13	52.0	363	21	AAA50159	Antibody 5H7 heavy
c 71		13	52.0	378	13	AAQ25667	Sequence of the an
72		13	52.0	387	21	AAC79285	Human lung tumour-
c 73		13	52.0	408	21	AAA74604	Murine 5B3 antibod
74		13	52.0	409	21	AAC03836	Human secreted pro
c 75		13	52.0	415	21	AAC56783	Eucalyptus grandis
c 76		13	52.0	441	21	AAC07649	Human secreted pro
77		13	52.0	476	21	AAC07951	Human secreted pro
c 78		13	52.0	476	21	AAC09392	Human secreted pro
79		13	52.0	498	22	AAS05926	Human cyclophilin
c 80		13	52.0	513	22	AAI26628	Probe #16561 for g
c 81		13	52.0	513	22	AAI55392	Probe #24078 used
c 82		13	52.0	516	22	AAI20799	Probe #10732 for g
c 83		13	52.0	516	22	AAI46029	Probe #14715 used
c 84		13	52.0	516	22	AAI06507	Probe #6498 used t

85	13	52.0	520	22	AAI18076	Probe #8009 for ge
86	13	52.0	520	22	AAI43076	Probe #11762 used
87	13	52.0	528	22	AAI18231	Probe #8164 for ge
88	13	52.0	528	22	AAI43274	Probe #11960 used
c 89	13	52.0	550	21	AAC94878	Cat flea hindgut a
90	13	52.0	556	22	AAH09733	Human cDNA clone (
91	13	52.0	556	22	AAH10417	Human cDNA clone (
c 92	13	52.0	559	22	AAF93531	cDNA encoding SRT
93	13	52.0	566	21	AAC07952	Human secreted pro
94	13	52.0	580	22	AAI16719	Probe #6652 for ge
95	13	52.0	580	22	AAI40105	Probe #8791 used t
c 96	13	52.0	585	22	AAI17412	Probe #7345 for ge
c 97	13	52.0	585	22	AAI42313	Probe #10999 used
98	13	52.0	603	22	AAH32562	Human secreted pro
c 99	13	52.0	607	21	AAA16367	Human colon cancer
100	13	52.0	626	21	AAA16245	Human colon cancer

ALIGNMENTS

RESULT 1
AAI62941
ID AAI62941 standard; DNA; 839 BP.
XX
AC AAI62941;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human genomic DNA SEQ ID NO 269.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200155449-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01346.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0218290.
PR 14-AUG-2000; 2000US-0225447.
PR 01-SEP-2000; 2000US-0229343.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231243.
PR 25-SEP-2000; 2000US-0234997.
PR 29-SEP-2000; 2000US-0236367.
PR 13-OCT-2000; 2000US-0239937.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246528.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249265.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476225/51.
XX
PT Novel plasma membrane associated proteins useful for diagnosing,
PT treating, preventing and/or prognosing disorders related to the
PT proteins, including cancer, immune response and neuronal disorders -
XX
PS Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;
Best Local Similarity 100.0%; Pred. No. 1.8e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 attgtcgcaagtcacagaattattt 25
||||||||||||||||||||||||||
Db 56 attgtcgcaagtcacagaattattt 80

RESULT 2
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX

```
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Patton DA;
XX
DR   WPI; 1999-152902/13.
DR   P-PSDB; AAW73906.
XX
PT   Transgenic plants with high biotin levels - transformed with DNA
PT   encoding di:amino-pelargonic acid amino-transferase or biotin
PT   synthase
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     24 attgtcgcaagtcacagaattattt 48


RESULT   3
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     47 attgtcgcaagtcacagaattattt 71


RESULT   4
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcacagaattattt 25
          |||||||||||||||||||||||||
Db     65 attgtcgcaagtcacagaattattt 89


RESULT   5
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
```

```
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db     140 attgtcgcaagtcacagaattattt 164


RESULT   6
AAQ56307/c
ID   AAQ56307 standard; DNA; 17 BP.
XX
AC   AAQ56307;
```

XX
DT 13-APR-1994 (first entry)
XX
DE BioB DNA primer.
XX
KW BioA; BioB; promoter; biotin; operon; primer; ss.
XX
OS Synthetic.
XX
PN JP05219956-A.
XX
PD 31-AUG-1993.
XX
PF 14-SEP-1992; 92JP-0244792.
XX
PR 14-SEP-1992; 92JP-0244792.
XX
PA (SHIS) SHISEIDO CO LTD.
PA (TAKE) TAKEDA CHEM IND LTD.
XX
DR WPI; 1993-308323/39.
XX
PT DNA sequence of biotin operon - has base sequence of E. coli
PT mutated by base pair(s) compared to wild type
XX
PS Example 1; Fig 8; 11pp; Japanese.
XX
CC A novel DNA sequence comprises the E.coli biotin operon (BO) in which
CC the control region of BO or the region near the bioB initiation
CC codon is mutated by at least one base pair compared to its
CC wild type. Two primers (AAQ56306-Q56307) are described in Example 1.
CC A microorganism belonging to Escherichia genus, transformed by
CC a recombinant plasmid carrying such DNA can be used for the prodn.
CC of biotin-active substances.
XX
SQ Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;

Query Match 68.0%; Score 17; DB 14; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 gtcgcaagtcacagaat 20
||||||||||||||||
Db 17 GTCGCAAGTCACAGAAT 1

RESULT 7
AAF94235
ID AAF94235 standard; DNA; 561 BP.
XX
AC AAF94235;
XX
DT 23-MAY-2001 (first entry)
XX
DE Primer specific for DNA encoding secretory/membrane protein SEQ ID 669.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; PCR primer; ss.
XX
OS Synthetic.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR WPI; 2001-093989/11.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
PS Claim 5; SEQ ID 669; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
SQ Sequence 561 BP; 213 A; 83 C; 114 G; 149 T; 2 other;

Query Match 60.0%; Score 15; DB 22; Length 561;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacagaatta 22
|||||||||||||||
Db 339 caagtcacagaatta 353

RESULT 8
AAZ30602
ID AAZ30602 standard; DNA; 1771 BP.
XX
AC AAZ30602;
XX
DT 18-JAN-2000 (first entry)
XX
DE Mouse integrin alpha 4 coding sequence.
XX
KW Human; integrin; antisense; oligonucleotide; inhibition; expression;
KW very late antigen; CD49d; CD29; cell surface; leucocyte; adhesion;
KW vascular endothelial cell; vascular endothelium; migration; inflammation;
KW atherosclerosis; allergy; asthma; rheumatoid arthritis; tumor;
KW metastasis; circulatory system; autoimmune disease; Grave's disease;
KW Hashimoto's thyroiditis; encephalomyelitis; multiple sclerosis; ds.
XX
OS Mus sp.
XX
PN US5968826-A.
XX
PD 19-OCT-1999.
XX
PF 05-OCT-1998; 98US-0166203.
XX
PR 05-OCT-1998; 98US-0166203.

```
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Bennett CF,  Cowsert LM,  Condon TP;
XX
DR   WPI; 1999-590416/50.
DR   P-PSDB; AAY44154.
XX
PT   Antisense inhibition of integrin alpha4 expression useful for treating
PT   inflammatory diseases such as atherosclerosis, allergies, asthma and
PT   arthritis -
XX
PS   Example 12; Column 71-74; 40pp; English.
XX
CC   This sequence represents the coding region of the mouse integrin
CC   alpha4 gene. The invention relates to the generation of antisense
CC   oligonucleotides targeted to the integrin alpha4 gene which are used
CC   for inhibiting expression of the integrin alpha4 mRNA or protein.
CC   Integrin alpha4 is a component of Very Late Antigen (VLA)-4 (also
CC   called alpha4beta1 and CD49d/CD29). VLA-4 is expressed on the cell
CC   surfaces of leucocytes and vascular endothelial cells and mediates the
CC   adhesion of leucocytes to the vascular endothelium prior to migration
CC   into the surrounding tissues. This migration is an essential step in
CC   inflammation and hence VLA-4 (and consequently integrin alpha4) is a
CC   potential therapeutic target for treating inflammatory diseases and
CC   the damaging effects of excessive inflammation. These disorders include
CC   atherosclerosis, allergies, asthma, rheumatoid arthritis and tumor cell
CC   metastasis (VLA-4 is involved in migration of the tumor cells through
CC   the extracellular matrix into the circulatory system). VLA-4 is also
CC   involved in a number of autoimmune diseases such as Grave's disease,
CC   Hashimoto's thyroiditis, encephalomyelitis (EAE), multiple sclerosis.
CC   VLA-4 may also be involved in promoting adhesion (i.e. retention) of
CC   hemopoietic stem cells in bone-marrow and in allograft rejection.
XX
SQ   Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 20;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaatt 21
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT   9
AAS10264
ID   AAS10264 standard; DNA; 1771 BP.
XX
AC   AAS10264;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Mouse integrin alpha 4 partial gene sequence.
XX
KW   Mouse; integrin alpha 4; antisense; L20788; very late antigen 4; VLA4;
KW   autoimmune disease; inflammatory disease; rheumatoid arthritis;
KW   multiple sclerosis; tumour metastasis; melanoma; asthma; psoriasis;
KW   allergy; Grave's disease; Hashimoto's thyroiditis;
KW   systemic lupus erythematosus; allograft rejection; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1193..1771
FT                   /*tag=  a
FT                   /product= "Integrin alpha 4"
FT                   /partial
FT                   /note= "No stop codon"
FT   intron          1387..1708
FT                   /*tag=  b
FT                   /number= 1
XX
PN   US6258790-B1.
XX
PD   10-JUL-2001.
XX
PF   19-AUG-1999;   99US-0377309.
XX
PR   05-OCT-1998;   98US-0166203.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Bennett CF,  Condon TP,  Cowsert LM;
XX
DR   WPI; 2001-450381/48.
DR   P-PSDB; AAU04967.
XX
PT   Composition for treating inflammatory and autoimmune diseases,
PT   comprises antisense compound targeted to nucleic acid molecule encoding
PT   integrin alpha4 and inhibit expression of integrin alpha4 -
XX
PS   Example 12; Column 77-79; 49pp; English.
XX
CC   The sequence (Genbank accession number L20788) is the partial gene
CC   sequence of mouse integrin alpha 4, one component of very late antigen 4
CC   (VLA4) which is involved in various autoimmune and inflammatory diseases.
CC   The invention relates to antisense inhibitors of integrin alpha 4 which
CC   target and inhibit expression of integrin alpha 4. The antisense
CC   molecules are useful for inhibiting the expression of integrin alpha4 in
CC   human cells or tissues, treating an animal having a disease or
CC   condition associated with expression of integrin alpha4, e.g.,
CC   inflammatory disease or condition, autoimmune disease or condition
CC   including rheumatoid arthritis, multiple sclerosis and tumour
CC   metastases, melanoma, asthma, psoriasis, allergy, Grave's disease,
CC   Hashimoto's thyroiditis, systemic lupus erythematosus and allograft
CC   rejection, and diseases or conditions characterised by leukocyte
CC   migration into affected tissues, preferably central nervous system
CC   tissues. The antisense molecules are also useful for reducing the
CC   levels of VLA-4 and alpha4beta7 integrin in human cells or tissues,
CC   and reducing the adherence of cells of a first type e.g., melanoma cells
CC   or lymphocytes, to cells of a second type e.g., endothelial cells, by
CC   inhibiting integrin alpha4 expression and thus decreasing adhesion of
CC   cells.
XX
SQ   Sequence 1771 BP; 371 A; 471 C; 519 G; 410 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 22;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaatt 21
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT  10
AAH18561/c
ID   AAH18561 standard; cDNA; 3210 BP.
XX
AC   AAH18561;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:18732.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
```

```
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 18732; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 3210 BP; 876 A; 753 C; 607 G; 974 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 22;  Length 3210;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 gtcacagaattattt 25
           |||||||||||||||
Db    2201 GTCACAGAATTATTT 2187


RESULT  11
AAF93909/c
ID   AAF93909 standard; cDNA; 3520 BP.
XX
AC   AAF93909;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Human cDNA encoding a membrane or secretory protein clone PSEC0256.
XX
KW   Human; secretory protein; membrane protein; vaccine; gene therapy;
KW   rheumatoid arthritis; diabetes; ss.
XX
OS   Homo sapiens.
XX
PN   EP1067182-A2.
XX
PD   10-JAN-2001.
XX
PF   07-JUL-2000; 2000EP-0114090.
XX
PR   08-JUL-1999;   99JP-0194179.
PR   11-JAN-2000; 2000JP-0118775.
PR   02-MAY-2000; 2000JP-0183766.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Kawai Y,  Sugiyama T,  Hayashi K;
XX
DR   WPI; 2001-093989/11.
DR   P-PSDB; AAB88482.
XX
PT   Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT   gene therapy or as candidate target molecules in drug development -
XX
PS   Claim 1; SEQ ID 331; 609pp + CD ROM; English.
XX
CC   This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC   which encode human secretory or membrane proteins represented by
CC   AAB88317 - AAB88419. Included in the invention are primers
CC   AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC   cDNA sequences of the invention. The invention also includes methods for
CC   the production of antibodies directed against the proteins, and cDNA
CC   sequences, which can be used in vaccines. The polynucleotide sequences
CC   can be used in gene therapy. The polynucleotide sequences and the
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with inappropriate secretory
CC   protein/membrane protein expression. The nucleic acids and complementary
CC   sequences may also be used as DNA probes in diagnostic assays
CC   (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC   presence of similar nucleic acid sequences in samples. They may also be
CC   used to study the expression and function of secretory proteins/membrane
CC   polypeptides and their role in metabolism. The polypeptides may be used
CC   as antigens in the production of antibodies against them and in assays to
CC   identify modulators (agonists and antagonists) of expression and
CC   activity. The antibodies and antagonists may also be used as therapeutic
CC   agents to down regulate expression and activity. The antibodies may also
CC   be used as diagnostic agents for detecting the presence of the
CC   polypeptides in samples (e.g. by enzyme linked immunosorbant assay
CC   (ELISA). Examples of diseases which may be treated include rheumatoid
CC   arthritis and diabetes.
XX
SQ   Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 22;  Length 3520;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           |||||||||||||||
Db    3182 CAAGTCACAGAATTA 3168


RESULT  12
AAF29361/c
ID   AAF29361 standard; DNA; 3520 BP.
XX
AC   AAF29361;
XX
DT   20-APR-2001  (first entry)
XX
DE   Amyloid-beta protein agglutination regulating factor DNA SEQ ID 9.
XX
KW   Human; amyloid-beta protein; agglutination regulatory factor;
KW   Alzheimer's disease; ds.
XX
OS   Homo sapiens.
XX
```

PN WO200104299-A1.
XX
PD 18-JAN-2001.
XX
PF 06-JUL-2000; 2000WO-JP04515.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 18-OCT-1999; 99US-0159586.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Yamazaki M, Satoh S;
PI Arakawa H, Morita M;
XX
DR WPI; 2001-138347/14.
DR P-PSDB; AAB49771.
XX
PT Polynucleotide encoding Amyloid-beta protein agglutination-controlling
PT factor, useful for inhibiting or promoting agglutination or
PT sedimentation of amyloid-beta protein and in diagnosis and screening
PT drugs for Alzheimer's disease -
XX
PS Claim 1; Page 58-63; 72pp; Japanese.
XX
CC This invention relates to polynucleotides AAF29357 - AAF29361 which
CC encode proteins AAB49767 - AAB49771. The proteins inhibit or promote the
CC agglutination of amyloid beta protein. The protein and polynucleotide
CC sequences are useful in the diagnosis of Alzheimer's disease. They are
CC also useful for screening drugs which are useful for treating Alzheimer's
CC disease.
XX
SQ Sequence 3520 BP; 871 A; 896 C; 901 G; 852 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 3520;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacagaatta 22
||||||||||||||||
Db 3182 CAAGTCACAGAATTA 3168

RESULT 13
AAF98729/c
ID AAF98729 standard; DNA; 3748 BP.
XX
AC AAF98729;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human late stage ovarian tumour polynucleotide marker 44.
XX
KW Human; ovarian cancer; identification; detection; characterisation;
KW tumour; kinase; marker; cytostatic; antisense gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200118542-A2.
XX
PD 15-MAR-2001.
XX
PF 01-SEP-2000; 2000WO-US24199.
XX
PR 03-SEP-1999; 99US-0152547.
PR 16-MAR-2000; 2000US-0190347.
PR 21-MAR-2000; 2000US-0191321.
PR 31-MAY-2000; 2000US-0208382.
PR 20-JUL-2000; 2000US-0220467.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lee J, Thompsho P, Lillie J;
XX
DR WPI; 2001-211428/21.
XX
PT Detection, assessment, prevention and therapy of ovarian cancer,
PT comprises detecting changes in the expression of a variety of markers -
XX
PS Claim 1; Page 1196-1197; 1198pp; English.
XX
CC The present invention describes a method for assessing whether a patient
CC is afflicted with ovarian cancer by comparing: (1) the expression of a
CC marker (I) (see AAF98594 to AAF98730), in a patient sample; and (2) the
CC normal level of expression of (I) in a control non-ovarian cancer
CC sample, where a significant difference between the level of expression
CC in (a) and (b) is an indication that the patient is afflicted with
CC ovarian cancer. (I) have cytostatic activities and can be used in
CC antisense gene therapy. The method, compositions and kits from the
CC present invention can be used for: (1) assessing and treating ovarian
CC cancer; (2) making isolated hybridoma, which produces an antibody useful
CC for ovarian cancer assessment; and (3) inhibiting ovarian cancer in a
CC patient. AAF98573 to AAF98593 represent human kinase marker primers and
CC probes which are used in the exemplification of the present invention.
XX
SQ Sequence 3748 BP; 1110 A; 825 C; 761 G; 1052 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 3748;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacagaatta 22
||||||||||||||||
Db 2627 CAAGTCACAGAATTA 2613

RESULT 14
AAS22683
ID AAS22683 standard; cDNA; 6087 BP.
XX
AC AAS22683;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human cDNA encoding a novel human protein #249.
XX
KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
PN WO200155437-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02623.
XX
PR 25-JAN-2000; 2000US-0491404.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-451939/48.
DR P-PSDB; AAU14378.
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX
PS Claim 1; Page 676-677; 894pp; English.

XX
CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amylotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence encodes a protein of the invention.
XX
SQ Sequence 6087 BP; 1490 A; 1609 C; 1574 G; 1413 T; 1 other;

Query Match 60.0%; Score 15; DB 22; Length 6087;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 tcgcaagtcacagaa 19
||||||||||||||||
Db 5191 tcgcaagtcacagaa 5205

RESULT 15
AAS22447
ID AAS22447 standard; cDNA; 6116 BP.
XX
AC AAS22447;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human cDNA encoding a novel human protein #13.
XX
KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
PN WO200155437-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02623.
XX
PR 25-JAN-2000; 2000US-0491404.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-451939/48.
DR P-PSDB; AAU14142.
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX
PS Claim 1; Page 190-196; 894pp; English.
XX
CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amylotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence encodes a protein of the invention.
XX
SQ Sequence 6116 BP; 1491 A; 1626 C; 1582 G; 1417 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 6116;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 tcgcaagtcacagaa 19
||||||||||||||||
Db 5220 tcgcaagtcacagaa 5234

RESULT 16
AAV52850/c
ID AAV52850 standard; DNA; 116624 BP.
XX
AC AAV52850;
XX
DT 06-NOV-1998 (first entry)
XX
DE Human eya1 gene contig 4405-9480.
XX
KW Human; eya1; branchio-oto-renal syndrome; BOR; carcinogenesis; cancer;
KW kidney tumourogenesis; tumour; diagnosis; ss.
XX
OS Homo sapiens.
XX
PN WO9832849-A2.
XX
PD 30-JUL-1998.
XX
PF 28-JAN-1998; 98WO-EP00433.
XX
PR 29-JAN-1997; 97US-0036579.
XX
PA (INSP) INST PASTEUR.
XX
PI Abdelhak S, Compain S, Petit C, Vasiliki K, Vincent C;
PI Weil D;
XX
DR WPI; 1998-427945/36.

```
XX
PT   Nucleic acid corresponding to human genes implicated in
PT   branchio-oto-renal syndrome - useful for, e.g. diagnosis and
PT   treatment of syndrome and possibly some carcinogenic processes,
PT   particularly in kidneys
XX
PS   Claim 1; Page 44-105; 191pp; English.
XX
CC   The present sequence represents a portion of the human eya-1 gene.
CC   Alterations in the eyal gene are associated with branchio-oto-renal
CC   (BOR) syndrome, including the form (BO) without renal anomalies, and
CC   possibly in some carcinogenic processes, particularly in the kidney.
CC   The related genes eya-2 and eya-3 may also be involved in development
CC   of tumours and cell differentiation. Antibodies specific for EYA
CC   proteins are used to detect the proteins by immunoassay, while genetic
CC   alterations linked to BOR are detected by amplifying DNA with primers
CC   (see AAV52857 to AAV52893) then identifying any mutations or deletions.
CC   EYA1 and EYA1-B proteins, corresponding nucleic acid, antisense sequences
CC   and Ab can be used therapeutically to modulate expression of EYA1(B), or
CC   their active derivatives, especially for treating renal disease
CC   associated with eyal abnormalities, also for ensuring correct
CC   development of grafted organs, especially where these are embryonic.
CC   Therapeutic proteins are administered to maintain 0.1-10 mu g /ml in
CC   body fluids, locally or systemically. Nucleic acid is administered in
CC   usual gene therapy vectors (optionally to cells ex vivo) at doses of
CC   0.1-100 mu g.
XX
SQ   Sequence 116624 BP; 35549 A; 20868 C; 23043 G; 37152 T; 12 other;


  Query Match             60.0%;  Score 15;  DB 19;  Length 116624;
  Best Local Similarity   100.0%;  Pred. No. 9.2;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 gtcacagaattattt 25
           |||||||||||||||
Db   19735 GTCACAGAATTATTT 19721


RESULT  17
AAA41925
ID   AAA41925 standard; cDNA; 329 BP.
XX
AC   AAA41925;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human secreted expressed sequence tag SEQ ID NO:665.
XX
KW   Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW   vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW   tumour; infection; depression; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200021990-A1.
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;   99WO-US24205.
XX
PR   15-OCT-1998;   98US-0104435.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M;
XX
DR   WPI; 2000-317937/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (sESTs), useful for treating various disorders
PT   such as autoimmune, infectious, and central nervous system disorders -
XX
PS   Claim 1; Page 308; 618pp; English.
XX
CC   AAA41261 to AAA43419 represent specifically claimed secreted expressed
CC   sequence tags (sESTs), isolated from human, mouse, xenopus and rat
CC   tissue sources. The sESTs can have a range of activities depending on
CC   the tissues they were isolated from. The activities include:
CC   chemotactic; proliferative; immunomodulatory; haematopoietic;
CC   chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC   cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC   antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC   nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used for gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC   disease), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA43420 to AAA43425 represent linker variants which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 329 BP; 97 A; 84 C; 67 G; 81 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 329;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattattt 25
           ||||||||||||||
Db    198 tcacagaattattt 211


RESULT  18
AAF65421
ID   AAF65421 standard; cDNA; 360 BP.
XX
AC   AAF65421;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 1177.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
```

```
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Lamson G,  Drmanac R;
PI   Crkenjakov R,  Drmanac S,  Dickson M,  Labat I,  Leshkowitiz D;
PI   Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 712; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 360 BP; 95 A; 70 C; 69 G; 125 T; 1 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 gtcacagaattatt 24
           ||||||||||||||
Db      71 gtcacagaattatt 84


RESULT  19
AAQ39649
ID   AAQ39649 standard; DNA; 385 BP.
XX
AC   AAQ39649;
XX
DT   20-MAY-1993  (first entry)
XX
DE   Expressed Sequence Tag human gene marker EST00025.
XX
KW   expressed sequence tag; human genome project; chromosome;
KW   human gene sequencing; PCR mapping; somatic cell hybrids;
KW   sublocalisation; gene tagging; tissue typing.
XX
OS   Synthetic.
XX
PN   WO9300353-A.
XX
PD   07-JAN-1993.
XX
PF   19-JUN-1992;   92WO-US05222.
XX
PR   20-JUN-1991;   91US-0716831.
PR   12-FEB-1992;   92US-0837195.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Adams MD,  Venter JC;
XX
DR   WPI; 1993-036325/04.
XX
PT   Particular expressed sequence tags from human CDNA - corresponds
PT   to transcription prods. of genes, useful for tagging genes,
PT   mapping chromosomes and tissue typing
XX
PS   Claim 3; Page 86; 199pp; English.
XX
CC   This sequence represents an EST (expressed sequence tag) ESTs are
CC   markers for human genes actually transcribed in vivo. Unlike the random
CC   genomic DNA sequence tagged sites (STSs), ESTs point directly to
CC   expressed genes. The use of ESTs could facilitate the tagging of most
CC   expressed human genes within a few years at a fraction of the cost of
CC   complete genomic sequencing. Using PCR primers AAQ39419-Q39580 (sequences
CC   designed from the ESTs) sublocalisation of an EST can be achieved with
CC   panels of fragments from specific chromosomes or pools of large genomic
CC   clones in an analogous manner. This sequence represents EST00025.
XX
SQ   Sequence 385 BP; 101 A; 66 C; 78 G; 139 T; 1 other;


  Query Match             56.0%;  Score 14;  DB 14;  Length 385;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattattt 25
           ||||||||||||||
Db     116 tcacagaattattt 129


RESULT  20
AAQ59061
ID   AAQ59061 standard; cDNA; 385 BP.
XX
AC   AAQ59061;
XX
DT   16-MAR-1994  (first entry)
XX
DE   Human brain Expressed Sequence Tag EST00025.
XX
KW   Gene transcription product; genetic markers; tagging; in vivo;
KW   transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS   Homo sapiens.
XX
PN   WO9316178-A.
XX
PD   19-AUG-1993.
XX
PF   12-FEB-1993;   93WO-US01294.
XX
PR   12-FEB-1992;   92US-0837195.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Adams MD,  Moreno RF,  Venter CJ;
XX
DR   WPI; 1993-272882/34.
XX
PT   Enriched oligonucleotides and corresp. sequences - used as
PT   markers for human genes transcribed in-vivo, facilitate tagging
PT   of most human genes
XX
PS   Example 1; Page 115; 500pp; English.
XX
CC   The Expressed Sequence Tag was isolated from a human brain cDNA
CC   library as part of a large set of ESTs which can be used as markers
CC   for human genes transcribed in vivo. They can be used to facilitate
CC   tagging of most human genes, for mapping locations of expressed genes
CC   on chromosomes, for individual or forensic identification, for mapping
CC   locations of disease-associated genes, for identification of tissue
CC   type, and for prepn. of antisense sequences, probes and constructs.
```

```
CC   EST00025 has a "poor" coding probability as evaluated using the
CC   coding-region prediction program CRM and hybridises to chromosome 20.
CC   See also AAQ59041-Q61440.
XX
SQ   Sequence 385 BP; 99 A; 66 C; 78 G; 140 T; 2 other;


  Query Match             56.0%;  Score 14;  DB 14;  Length 385;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattattt 25
           |||||||||||||||
Db     116 tcacagaattattt 129


RESULT  21
AAI13743
ID   AAI13743 standard; DNA; 403 BP.
XX
AC   AAI13743;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #3676 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 3676; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 403 BP; 107 A; 87 C; 91 G; 118 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 403;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattattt 25
           |||||||||||||||
Db     162 tcacagaattattt 175


RESULT  22
AAI35103
ID   AAI35103 standard; DNA; 403 BP.
XX
AC   AAI35103;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #3789 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 3789; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 403 BP; 107 A; 87 C; 91 G; 118 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 403;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattattt 25
           |||||||||||||||
Db     162 tcacagaattattt 175


RESULT  23
AAI03624
ID   AAI03624 standard; DNA; 403 BP.
XX
AC   AAI03624;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Probe #3615 used to measure gene expression in human breast sample.
```

```
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 3615; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 403 BP; 107 A; 87 C; 91 G; 118 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 403;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattattt 25
           ||||||||||||||
Db     162 tcacagaattattt 175


RESULT  24
AAH12853
ID   AAH12853 standard; cDNA; 523 BP.
XX
AC   AAH12853;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:9688.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 3; SEQ ID 9688; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 523 BP; 206 A; 87 C; 70 G; 151 T; 9 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 523;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatt 21
           ||||||||||||||
Db     390 caagtcacagaatt 403


RESULT  25
AAI41094/c
ID   AAI41094 standard; DNA; 574 BP.
XX
AC   AAI41094;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #9780 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
```

XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 9780; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 574 BP; 229 A; 109 C; 104 G; 132 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 574;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 tcacagaattattt 25
||||||||||||||
Db 48 TCACAGAATTATTT 35

RESULT 26
AAZ56756/c
ID AAZ56756 standard; cDNA; 902 BP.
XX
AC AAZ56756;
XX
DT 23-MAR-2000 (first entry)
XX
DE Human transmembrane protein HTMPN-59 encoding cDNA.
XX
KW Human; transmembrane protein; HTMPN; diagnosis; immunospecific;
KW antiproliferative; neuroprotective; immune disorder;
KW reproductive disorder; smooth muscle disorder; neurological disorder;
KW gastrointestinal disorder; developmental disorder;
KW cell proliferative disorder; ss.
XX
OS Homo sapiens.
XX
PN WO9961471-A2.
XX
PD 02-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11904.
XX
PR 29-MAY-1998; 98US-0087260.
PR 02-JUL-1998; 98US-0091674.
PR 02-OCT-1998; 98US-0102954.
PR 24-NOV-1998; 98US-0109869.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Tang YT, Lal P, Hillman JL, Yue H, Guegler KJ, Corley NC;
PI Bandman O, Patterson C, Gorgone GA, Kaser MR, Baughn MR;
PI Au-Young J;
XX
DR WPI; 2000-072605/06.
DR P-PSDB; AAY57935.
XX
PT Proteins, polynucleotides, vectors, host cells and antibodies used to
PT diagnose, treat or prevent immune, reproductive, smooth muscle,
PT neurological, gastrointestinal, developmental and cell proliferative
PT disorders -
XX
PS Claim 9; Page 215; 229pp; English.
XX
CC AAZ56698 to AAZ56776 encode AAY57877 to AAY57955 which represent human
CC transmembrane proteins designated HTMPN-1 to HTMPN-79, respectively.
CC The transmembrane protein have immunospecific, antiproliferative and
CC neuroprotective activities. The human transmembrane proteins,
CC polynucleotides encoding them and other compositions and methods from
CC the present invention, can be used for the diagnosis, treatment or
CC prevention of immune, reproductive, smooth muscle, neurological,
CC gastrointestinal, developmental and cell proliferative disorders. The
CC HTMPN's can be used to treat or prevent disorders associated with a
CC decreased expression or activity of HTMPN.
XX
SQ Sequence 902 BP; 261 A; 212 C; 196 G; 233 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 902;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 agtcacagaattat 23
||||||||||||||
Db 425 AGTCACAGAATTAT 412

RESULT 27
AAN80927/c
ID AAN80927 standard; DNA; 1050 BP.
XX
AC AAN80927;
XX
DT 15-OCT-1990 (first entry)
XX
DE Lividans Exported Protein gene.
XX
KW Longisporus trypsin inhibitor; Lividans exported protein; LTI; LEP-10;
KW protease inhibitor; probe; ss.
XX
OS Streptomyces lividans strain 1326.
XX
FH Key Location/Qualifiers
FT sig_peptide 394..498
FT /*tag= a
FT mat_peptide 499.. 825
FT /*tag= b
FT /product=LEP-10
XX
PN WO8801278-A.
XX
PD 25-FEB-1988.
XX
PF 17-AUG-1987; 87WO-US02009.
XX
PR 18-AUG-1986; 86US-0897245.
XX
PA (SMIK) SMITHKLINE BECKMAN CORP.
PA (BERK/) TR BERKA.
XX
PI Berka TR, Fornwald JA, Gorniak JG, Rosenberg M;

XX
DR WPI; 1988-063996/09.
DR P-PSDB; AAP80944.
XX
PT Novel protease inhibitors LEP-10 and LTI - produced by Streptomyces spp.
XX
PS Disclosure; ; pp; English.
XX
CC The sequence can be cloned and used to produce recombinant LEP-10.
CC The DNA can be used as a probe for identifying further coding
CC sequences for protease inhibitors; the protein is useful as an
CC assay reagent.
CC See also AAN80926.
XX
SQ Sequence 1050 BP; 149 A; 398 C; 366 G; 137 T; 0 other;

Query Match 56.0%; Score 14; DB 9; Length 1050;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 aagtcacagaatta 22
||||||||||||||
Db 251 AAGTCACAGAATTA 238

RESULT 28
AAH14046/c
ID AAH14046 standard; cDNA; 1465 BP.
XX
AC AAH14046;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:11168.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11168; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1465 BP; 470 A; 259 C; 270 G; 466 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1465;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacagaatt 21
||||||||||||||
Db 1076 CAAGTCACAGAATT 1063

RESULT 29
AAT31932/c
ID AAT31932 standard; DNA; 1677 BP.
XX
AC AAT31932;
XX
DT 27-SEP-1996 (first entry)
XX
DE Retinoid X receptor interacting protein RIP14-2 DNA.
XX
KW Retinoid X receptor interacting protein; RXR; RIP; RIP14-2; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..1356
FT /*tag= a
FT polyA_signal 1672-1677
FT /*tag= b
XX
PN WO9621677-A1.
XX
PD 18-JUL-1996.
XX
PF 08-DEC-1995; 95WO-US16311.
XX
PR 13-JAN-1995; 95US-0372652.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Choi H, Moore D, Seol W;
XX
DR WPI; 1996-342241/34.
DR P-PSDB; AAR99739.
XX
PT Retinoid X receptor (RXR) interacting protein (RIP) - useful to
PT modulate or mediate RXR function, anti-RIP antibodies can be used to
PT determine RIP subcellular distribution patterns
XX
PS Claim 8; Page 59-60; 90pp; English.
XX
CC A cDNA clone (AAT31932) codes for retinoid X receptor interacting
CC protein RIP14-2 (AAR99739), a previously undescribed orphan member
CC of the nuclear receptor superfamily which may be involved in the
CC complex retinoid response. It was isolated from a mouse liver

```
CC   cDNA library using an RIP14 probe.  Different RIP14-2 clones had
CC   different 5' sequences (see also AAT31935-37).  Another isoform,
CC   RIP14-1, was also detected (see also AAT31928).  The cDNA can be
CC   used for the recombinant prodn. of RIP14-2 in transformed host
CC   cells.
XX
SQ   Sequence 1677 BP; 486 A; 385 C; 408 G; 398 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 17;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgtcgcaagtcac 15
           ||||||||||||||
Db     630 TTGTCGCAAGTCAC 617


RESULT  30
AAT31928/c
ID   AAT31928 standard; DNA; 1787 BP.
XX
AC   AAT31928;
XX
DT   27-SEP-1996  (first entry)
XX
DE   Retinoid X receptor interacting protein RIP14-1 DNA.
XX
KW   Retinoid X receptor interacting protein; RXR; RIP; RIP14-1; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             12..1466
FT                   /*tag=  a
FT   polyA_signal    1782..1787
FT                   /*tag=  b
XX
PN   WO9621677-A1.
XX
PD   18-JUL-1996.
XX
PF   08-DEC-1995;   95WO-US16311.
XX
PR   13-JAN-1995;   95US-0372652.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Choi H,  Moore D,  Seol W;
XX
DR   WPI; 1996-342241/34.
DR   P-PSDB; AAR99735.
XX
PT   Retinoid X receptor (RXR) interacting protein (RIP)  - useful to
PT   modulate or mediate RXR function, anti-RIP antibodies can be used to
PT   determine RIP subcellular distribution patterns
XX
PS   Claim 8; Page 53-54; 90pp; English.
XX
CC   A full-length cDNA clone (AAT31928) codes for retinoid X receptor
CC   interacting protein RIP14-1 (AAR99735), a previously undescribed
CC   orphan member of the nuclear receptor superfamily which may be
CC   involved in the complex retinoid response.  It was isolated from
CC   a mouse liver cDNA library using an RIP14 probe.  Another isoform,
CC   RIP14-2, was also detected (see also AAT31932).  The cDNA can be used
CC   for the recombinant prodn. of RIP14-1 in transformed host cells.
XX
SQ   Sequence 1787 BP; 515 A; 408 C; 442 G; 422 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 17;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgtcgcaagtcac 15
           ||||||||||||||
Db     740 TTGTCGCAAGTCAC 727


RESULT  31
AAZ52940/c
ID   AAZ52940 standard; cDNA; 1890 BP.
XX
AC   AAZ52940;
XX
DT   14-MAR-2000  (first entry)
XX
DE   Human prostate tumor cDNA library derived EST fragment #83.
XX
KW   Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;
KW   treatment; ds.
XX
OS   Homo sapiens.
XX
PN   DE19820190-A1.
XX
PD   04-NOV-1999.
XX
PF   28-APR-1998;   98DE-1020190.
XX
PR   28-APR-1998;   98DE-1020190.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A,  Specht T,  Hinzmann B,  Schmitt A,  Pilarsky C,  Dahl E;
XX
DR   WPI; 1999-621386/54.
DR   P-PSDB; AAY74061, AAY74062, AAY74063.
XX
PT   New human nucleic acid sequences from pancreatic tumors, and related
PT   proteins  -
XX
PS   Claim 2; Page 250; 502pp; German.
XX
CC   This invention describes novel polypeptides and their encoding nucleic
CC   acids derived from human pancreatic tumor tissue which have cytostatic
CC   activity. The sequences are also useful in producing pharmaceutical
CC   compositions for treatment of pancreatic tumors. AAZ52858-Z53014
CC   represent expressed sequence tag (EST) fragments derived from a human
CC   pancreatic tumor cDNA library and which encode the proteins represented
CC   in AAY73814-Y74252.
XX
SQ   Sequence 1890 BP; 590 A; 358 C; 398 G; 544 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 1890;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatt 21
           ||||||||||||||
Db    1475 CAAGTCACAGAATT 1462


RESULT  32
AAI60385
ID   AAI60385 standard; cDNA; 1919 BP.
XX
AC   AAI60385;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4374.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
```

```
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM41229.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 4374; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 1919 BP; 564 A; 394 C; 343 G; 617 T; 1 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 1919;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatt 21
           ||||||||||||||
Db     402 caagtcacagaatt 415


RESULT  33
AAH15029/c
ID   AAH15029 standard; cDNA; 2022 BP.
XX
AC   AAH15029;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:13001.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 13001; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2022 BP; 622 A; 389 C; 418 G; 593 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 2022;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatt 21
           ||||||||||||||
Db    1633 CAAGTCACAGAATT 1620


RESULT  34
AAI58599/c
ID   AAI58599 standard; cDNA; 2102 BP.
XX
AC   AAI58599;
XX
```

DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 802.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM39443.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 802; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 2102 BP; 662 A; 390 C; 441 G; 609 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2102;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacagaatt 21
||||||||||||||
Db 1564 CAAGTCACAGAATT 1551

RESULT 35
AAD12786/c

ID AAD12786 standard; cDNA; 2220 BP.
XX
AC AAD12786;
XX
DT 23-OCT-2001 (first entry)
XX
DE Human nuclear hormone receptor (NHREC)-3 cDNA.
XX
KW Human; nuclear hormone receptor; NHREC-3; intracellular receptor; cancer;
KW transcriptional regulator; metabolic disorder; lysosomal storage disease;
KW Addison's disease; cystic fibrosis; endocrine disorder; atherosclerosis;
KW breast cancer; osteoporosis; Cushing's disease; reproductive disorder;
KW developmental disorder; infertility; renal tubular acidosis; cirrhosis;
KW endometriosis; hereditary neuropathy; immunological disorder; anaemia;
KW acquired immunodeficiency syndrome; AIDS; cell proliferative disorder;
KW allergy; Grave's disease; gene therapy; drug screening; cytostatic;
KW immunomodulatory; gout; diabetes; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 337..1770
FT /*tag= a
FT /product= "Human nuclear hormone receptor (NHREC)-3"
XX
PN WO200155392-A2.
XX
PD 02-AUG-2001.
XX
PF 26-JAN-2001; 2001WO-US02741.
XX
PR 28-JAN-2000; 2000US-0178578.
PR 25-FEB-2000; 2000US-0185079.
PR 17-MAR-2000; 2000US-0189999.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Burford N, Baughn MR, Tang YT, Bandman O, Tribouley CM, Das D;
PI Policky JL, Lo TP;
XX
DR WPI; 2001-488797/53.
DR P-PSDB; AAE06680.
XX
PT New human nuclear hormone receptors and polynucleotides encoding the
PT receptors, useful for diagnosing, treating and preventing metabolic,
PT endocrine, reproductive, developmental, immunological and cell
PT proliferative disorders -
XX
PS Claim 5; Page 95-96; 99pp; English.
XX
CC The present sequence is human nuclear hormone receptor (NHREC)-3 cDNA.
CC NHREC known as nuclear receptor or intracellular receptor, constitute
CC a protein superfamily whose members are both receptors and
CC transcriptional regulators. NHREC is useful in the diagnosis, treatment
CC and prevention of metabolic (e.g. Addison's disease, cystic fibrosis,
CC diabetes, lysosomal storage disease), endocrine (e.g. disorders of
CC pituitary, thyroid, Cushing's disease, osteoporosis), reproductive (e.g.
CC infertility, endometriosis, breast cancer), developmental (e.g. anaemia,
CC renal tubular acidosis, hereditary neuropathy), immunological disorders
CC (e.g. acquired immunodeficiency syndrome (AIDS), allergy, Grave's
CC disease, gout) and cell proliferative disorders (e.g. atherosclerosis,
CC cirrhosis, cancer) and in the assessment of the effects of exogenous
CC compounds on the expression of nucleic acid and amino acid sequences of
CC nuclear hormone receptors. NHREC may also be used to screen for
CC compounds that modulate the activity of NHREC, for somatic or germline
CC gene therapy, to detect single nucleotide polymorphisms, as element on
CC microarray, to generate hybridisation probes and in drug screening.
XX
SQ Sequence 2220 BP; 710 A; 425 C; 465 G; 620 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2220;
Best Local Similarity 100.0%; Pred. No. 35;

```
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db   1041 TTGTCGCAAGTCAC 1028


RESULT  36
AAV10120/c
ID   AAV10120 standard; cDNA; 2233 BP.
XX
AC   AAV10120;
XX
DT   29-MAY-1998  (first entry)
XX
DE   Human retinoid receptor RR1 cDNA.
XX
KW   Retinoid receptor; RR1; steroid receptor; agonist; antagonist; cancer;
KW   adrenal deficiency; skin disorder; inflammatory disorder;
KW   immune response regulator; autoimmune disease; therapeutic antibody; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             363..1781
FT                   /*tag=  a
FT                   /product= RR1
FT                   /note= "retinoid receptor"
XX
PN   US5728548-A.
XX
PD   17-MAR-1998.
XX
PF   29-JUN-1995;   95US-0496631.
XX
PR   29-JUN-1995;   95US-0496631.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Bowman M;
XX
DR   WPI; 1998-206567/18.
DR   P-PSDB; AAW40072.
XX
PT   Human retinoid receptor protein RR1 - useful for, e.g. drug
PT   screening, therapy and antibody production
XX
PS   Claim 8; Column 11-15; 13pp; English.
XX
CC   This sequence encodes a novel human steroid receptor, the retinoid
CC   receptor protein or RR1. This protein can be used in screening assays
CC   for steroid hormone receptor agonists and antagonists and in
CC   pharmaceutical compositions for treating adrenal deficiencies, e.g.
CC   Addison's disease, cancer, skin disorders, e.g. acne and psoriasis,
CC   inflammatory disorders, e.g. arthritis and HIV infections. The protein
CC   can also be used for regulating immune responses, e.g. as antitumour
CC   agents, vaccine adjuvants, organ rejection inhibitors or agents for
CC   treating autoimmune diseases. The protein can further be used to produce
CC   therapeutic antibodies.
XX
SQ   Sequence 2233 BP; 744 A; 427 C; 463 G; 599 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 19;  Length 2233;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db   1055 TTGTCGCAAGTCAC 1042


RESULT  37
AAF77541/c
ID   AAF77541 standard; DNA; 2585 BP.
XX
AC   AAF77541;
XX
DT   23-MAY-2001  (first entry)
XX
DE   Caenorhabditis briggsae mab-21 coding sequence.
XX
KW   Mab-21; cell fate choice; pax-6; aniridia; Moebius syndrome;
KW   chromosome 13q13; peripheral neural tissue differentiation; ds.
XX
OS   Caenorhabditis briggsae.
XX
PN   US6197504-B1.
XX
PD   06-MAR-2001.
XX
PF   19-JAN-1998;   98US-0008697.
XX
PR   10-APR-1996;   96US-0011607.
PR   09-APR-1997;   97US-0835604.
XX
PA   (CHOW/) CHOW K L.
XX
PI   Chow KL;
XX
DR   WPI; 2001-256362/26.
DR   P-PSDB; AAB74419.
XX
PT   Detecting expression of mab-21 gene of Caenorhabditis elegans encoding
PT   novel protein required for choice of alternate cell fates in sample, by
PT   contacting mRNA from sample with nucleic acid hybridizing with mab-21 -
XX
PS   Disclosure; column 41-44; 79pp; English.
XX
CC   The present invention describes a method of detecting mab-21 expression
CC   in a sample involving contacting total mRNA from the sample with a probe
CC   to the mab-21 gene. The sequences of the mab-21 coding sequence and
CC   protein from several species are also given. The human mab-21 gene is
CC   found on chromosome 13q13, a region which has been linked to Moebius
CC   syndrome. The mab-21 protein shows homology to pax-6, mutations in which
CC   are associated with aniridia and which is linked to peripheral neural
CC   tissue differentiation.
XX
SQ   Sequence 2585 BP; 751 A; 540 C; 559 G; 688 T; 47 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 2585;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatt 21
          ||||||||||||||
Db   1058 CAAGTCACAGAATT 1045


RESULT  38
AAC69462/c
ID   AAC69462 standard; cDNA; 2845 BP.
XX
AC   AAC69462;
XX
DT   30-JAN-2001  (first entry)
XX
DE   Human secreted protein gene 8 SEQ ID NO:18.
XX
KW   Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW   fungicide; ophthalmological; gene therapy; autoimmune disease; infection;
KW   hyperproliferative disorder; cardiovascular disorder; angiogenesis;
KW   cerebrovascular disorder; nervous system disorder; ocular disorder;
```

KW wound healing; skin aging; food additive; preservative; ss.
XX
OS Homo sapiens.
XX
PN WO200058469-A1.
XX
PD 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07579.
XX
PR 26-MAR-1999; 99US-0126509.
PR 07-JAN-2000; 2000US-0174853.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-594642/56.
DR P-PSDB; AAB38210.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 1; Page 334-335; 416pp; English.
XX
CC The polynucleotide sequences given in AAC69455 to AAC69502 encode the
CC human secreted proteins given in AAB38203 to AAB38250. AAB38251 to
CC AAB38320 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Example of activities include:
CC immunosuppressive; antiarthritic; antirheumatic; antiproliferative;
CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC neuroprotective; antibacterial; virucide; fungicide; and
CC ophthalmological. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases, hyperproliferative disorders, cardiovascular
CC disorders, cerebrovascular disorders, angiogenesis, nervous system
CC disorders, infections caused by bacteria, viruses and fungi and ocular
CC disorders. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. AAC69446 to AAC69454 and
CC AAB38202 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 2845 BP; 871 A; 518 C; 439 G; 1007 T; 10 other;

Query Match 56.0%; Score 14; DB 21; Length 2845;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 tcacagaattattt 25
||||||||||||||
Db 1681 TCACAGAATTATTT 1668

RESULT 39
AAV72330/c
ID AAV72330 standard; DNA; 3413 BP.
XX
AC AAV72330;
XX
DT 28-JUL-1999 (first entry)
XX
DE F. balustinum CP70 DNA.
XX
KW CP70; cold active protease; detergent; cold temperature wash water; ss.
XX
OS Flavobacterium balustinum.
XX
FH Key Location/Qualifiers
FT CDS 437..2536
FT /*tag= a
FT /product= "CP70"
XX
PN WO9925848-A1.
XX
PD 27-MAY-1999.
XX
PF 14-NOV-1997; 97WO-US20843.
XX
PR 14-NOV-1997; 97WO-US20843.
XX
PA (PROC) PROCTER & GAMBLE CO.
XX
PI Kitado H, Yoshikawa A;
XX
DR WPI; 1999-347483/29.
DR P-PSDB; AAY08471.
XX
PT Newly isolated CP70 polynucleotide exhibiting characteristics of
PT Deposit No. FERM BP-6154
XX
PS Claim 1a; Page 15-18; 25pp; English.
XX
CC This invention describes a novel cold active protease, CP70 from
CC Flavobacterium balustinum. The product of the invention is useful
CC in detergent compositions for use in cold temperature wash water.
CC The characterization of the polynucleotide facilitates the production
CC of an alternative relatively large source of cold active protease.
XX
SQ Sequence 3413 BP; 1228 A; 558 C; 586 G; 1041 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 3413;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 agtcacagaattat 23
||||||||||||||
Db 3107 AGTCACAGAATTAT 3094

RESULT 40
AAI61089/c
ID AAI61089 standard; cDNA; 5662 BP.
XX
AC AAI61089;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5078.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM41933.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 5078; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 5662 BP; 1631 A; 1126 C; 1247 G; 1658 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 5662;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 gcaagtcacagaat 20
||||||||||||||
Db 4497 GCAAGTCACAGAAT 4484

RESULT 41
AAC59953
ID AAC59953 standard; cDNA; 5684 BP.
XX
AC AAC59953;
XX
DT 30-JAN-2001 (first entry)
XX
DE Human secreted protein cDNA sequence #47.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200055198-A1.
XX
PD 21-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06012.
XX
PR 12-MAR-1999; 99US-0124093.
PR 23-NOV-1999; 99US-0166989.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI; 2000-587520/55.
DR P-PSDB; AAB28748.
XX
PT Human secreted proteins and the nucleic acids that encode them, useful
PT in gene therapy protocols and recombinant nucleic acid based procedures
PT -
XX
PS Claim 1; Page 343-344; 391pp; English.
XX
CC The invention relate to the isolation of genes AAC59907-C59956 encoding
CC 50 human secreted proteins AAB28702-B28751. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 5684 BP; 1641 A; 1250 C; 1045 G; 1732 T; 16 other;

Query Match 56.0%; Score 14; DB 21; Length 5684;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 tcacagaattattt 25
||||||||||||||
Db 4933 tcacagaattattt 4946

RESULT 42
AAC76014/c
ID AAC76014 standard; cDNA; 5773 BP.
XX
AC AAC76014;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1569 polynucleotide sequence SEQ ID NO:3137.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;

KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR P-PSDB; AAB41805.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 2354-2357; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;

Query Match 56.0%; Score 14; DB 21; Length 5773;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 gcaagtcacagaat 20
||||||||||||||
Db 4611 GCAAGTCACAGAAT 4598

RESULT 43
AAI59303/c
ID AAI59303 standard; cDNA; 5876 BP.
XX
AC AAI59303;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1506.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM40147.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 1506; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 5876 BP; 1704 A; 1165 C; 1288 G; 1719 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 5876;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 gcaagtcacagaat 20
||||||||||||||
Db 4739 GCAAGTCACAGAAT 4726

RESULT 44
AAH24651/c
ID AAH24651 standard; cDNA; 7063 BP.
XX
AC AAH24651;
XX
DT 06-AUG-2001 (first entry)

```
XX
DE   Nucleotide sequence of a disrupted in schizophrenia 1 (DIS1) gene.
XX
KW   Disrupted in schizophrenia 1 gene; DIS1 gene; schizophrenia;
KW   psychiatric disorder; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             54..2618
FT                   /*tag= a
FT                   /product= "disrupted in schizophrenia 1 (DIS1) protein"
XX
PN   WO200140301-A2.
XX
PD   07-JUN-2001.
XX
PF   28-NOV-2000; 2000WO-EP11915.
XX
PR   01-DEC-1999;   99EP-0309667.
XX
PA   (ALKU ) AKZO NOBEL NV.
PA   (MEDI-) MEDICAL RES COUNCIL.
PA   (UYED-) UNIV EDINBURGH.
XX
PI   Porteous D,  Millar K,  Blackwood D;
XX
DR   WPI; 2001-374796/39.
DR   P-PSDB; AAB84217.
XX
PT   Novel isolated polynucleotide which surrounds a breakpoint on
PT   chromosome 1 involved in a balanced t(1;11) (q42.1;q14.3)
PT   translocation, and its encoded proteins, useful as medicament for
PT   treating psychiatric disorders  -
XX
PS   Claim 2; Page 30-32; 51pp; English.
XX
CC   The present sequence encodes a human disrupted in schizophrenia 1 (DIS1)
CC   protein. DIS1 is disrupted by a (1:11)(q42.1:q14.3) translocation which
CC   segregates with schizophrenia. The DIS1 gene consists of 13 exons. Exon
CC   11 contains an alternative splice site which removed 66 nucleotides from
CC   the open reading frame. The final intron of DIS1 belongs to the rare
CC   AT-AC class of introns. DIS1 polypeptide fragments and antibodies are
CC   useful for in vitro diagnosis of a psychiatric disorder. DIS1 protein is
CC   useful in a screening assay for identifying new drugs for treating
CC   psychiatric disorders. DIS1 polypeptides and polynucleotides are useful
CC   for preventing and/or treating diseases related to schizophrenia.
XX
SQ   Sequence 7063 BP; 1930 A; 1644 C; 1690 G; 1799 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 7063;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattattt 25
           ||||||||||||||
Db    5149 TCACAGAATTATTT 5136


RESULT  45
AAT10105
ID   AAT10105 standard; DNA; 8041 BP.
XX
AC   AAT10105;
XX
DT   13-MAY-1996  (first entry)
XX
DE   Adherence conferring plasmid pear.
XX
KW   adhesin; plasmid pear; vector; vaccine; intestine colonisation; ds.
XX
OS   Chimeric Escherichia coli;
OS   Chimeric synthetic.
XX
FH   Key             Location/Qualifiers
FT   primer_bind     complement (2867..2890)
FT                   /*tag= a
FT                   /note= "primer for E.coli O157:H7 DNA"
FT   primer_bind     complement (2895..2914)
FT                   /*tag= b
FT                   /note= "primer for pear"
FT   CDS             3036..5126
FT                   /*tag= c
FT                   /product= adhesin
FT                   /trans_except=pos:3912..3914:_aa:Gly
FT                   /note= "claim 1, page 36"
FT   misc_feature    3271..3310
FT                   /*tag= d
FT                   /note= "region of transposon TnphoA insertion"
FT   misc_feature    3801..3840
FT                   /*tag= e
FT                   /note= "region of transposon TnPhoA insertion"
FT   primer_bind     complement (5176..5196)
FT                   /*tag= f
FT                   /note= "primer for pear"
FT   primer_bind     complement (5159..5183)
FT                   /*tag= g
FT                   /note= "primer for E.coli O157H7 DNA"
FT   CDS             complement (6449..7024)
FT                   /*tag= h
FT                   /note= "terE homologue"
FT   CDS             complement (7092..7670)
FT                   /*tag= i
FT                   /note= "terD homologue"
XX
PN   WO9600233-A1.
XX
PD   04-JAN-1996.
XX
PF   07-JUN-1995;   95WO-US06994.
XX
PR   24-JUN-1994;   94US-0265714.
XX
PA   (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
PA   (UNIW ) UNIV WASHINGTON.
PA   (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX
PI   Besser TE,  Bilge SS,  Tarr PI,  Vary JC;
XX
DR   WPI; 1996-068826/07.
DR   P-PSDB; AAR75366.
XX
PT   Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
PT   isolated on plasmid pSC (overlap), for use as a vaccine to prevent
PT   bacterial colonisation of bovine intestine
XX
PS   Disclosure; Page 23-30; 42pp; English.
XX
CC   Adherence conferring plasmid pear (AAT10105) comprises Escherichia
CC   coli O157:H7 chromosomal DNA plus the Stratagene SK+ vector.
CC   It includes 3 open reading frames, 2 of which are homologues of
CC   terE and terD genes necessary for tellurite resistance.  The
CC   third ORF is homologous to the IrgA gene and encodes an adhesin
CC   (AAR75366) that enables E. coli O157:H7, an antibiotic-resistant,
CC   virulent and common food-borne pathogen, to adhere to epithelial
CC   cells.  This ORF can be utilised in the prodn. of adhesin for use
CC   as a vaccine to prevent disease or colonisation of mucosal surfaces
CC   by O157:H7.
XX
SQ   Sequence 8041 BP; 2241 A; 1807 C; 1838 G; 2126 T; 29 other;


  Query Match             56.0%;  Score 14;  DB 17;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      10 agtcacagaattat 23
           ||||||||||||||
Db    3727 agtcacagaattat 3740
```

Search completed: December 26, 2001, 14:16:32
Job time: 6984 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:52:29 ; Search time 143.1 Seconds
(without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196F-2
Perfect score: 25
Sequence: 1 attgtcgcaagtcacagaattattt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 9

Total number of hits satisfying chosen parameters: 9633

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
	2	25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
	3	25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
	4	25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
c	5	17	68.0	17	2	US-07-944-259-2	Sequence 2, Appli
	6	15	60.0	1771	2	US-09-166-203-48	Sequence 48, Appl
	7	15	60.0	1771	4	US-09-377-309-48	Sequence 48, Appl
c	8	14	56.0	1050	6	5498529-7	Patent No. 5498529
c	9	14	56.0	1677	2	US-08-372-652-14	Sequence 14, Appl
c	10	14	56.0	1677	5	PCT-US95-16311-14	Sequence 14, Appl
c	11	14	56.0	1787	2	US-08-372-652-6	Sequence 6, Appli
c	12	14	56.0	1787	5	PCT-US95-16311-6	Sequence 6, Appli
	13	14	56.0	2091	1	US-08-765-081-4	Sequence 4, Appli
	14	14	56.0	2091	3	US-09-098-082-4	Sequence 4, Appli
	15	14	56.0	2091	5	PCT-US95-06994-4	Sequence 4, Appli
c	16	14	56.0	2233	1	US-08-496-631-1	Sequence 1, Appli
c	17	14	56.0	2585	4	US-09-008-697A-7	Sequence 7, Appli
	18	14	56.0	8041	1	US-08-765-081-1	Sequence 1, Appli
	19	14	56.0	8041	3	US-09-098-082-1	Sequence 1, Appli
	20	14	56.0	8041	5	PCT-US95-06994-1	Sequence 1, Appli
c	21	13	52.0	607	4	US-09-385-982-372	Sequence 372, App
	22	13	52.0	626	4	US-09-385-982-250	Sequence 250, App
	23	13	52.0	634	4	US-09-385-982-253	Sequence 253, App
	24	13	52.0	651	4	US-09-385-982-329	Sequence 329, App
c	25	13	52.0	1500	4	US-08-875-847B-3	Sequence 3, Appli
c	26	13	52.0	1505	4	US-08-875-847B-1	Sequence 1, Appli
c	27	13	52.0	1662	1	US-08-651-572-1	Sequence 1, Appli
c	28	13	52.0	1662	3	US-09-066-544-1	Sequence 1, Appli
c	29	13	52.0	1662	3	US-08-951-086-1	Sequence 1, Appli
c	30	13	52.0	2214	2	US-08-484-993B-1	Sequence 1, Appli
c	31	13	52.0	2214	2	US-08-484-158B-1	Sequence 1, Appli
c	32	13	52.0	2214	2	US-08-484-596A-1	Sequence 1, Appli
c	33	13	52.0	2214	2	US-08-480-150A-1	Sequence 1, Appli
c	34	13	52.0	2214	3	US-08-458-731-1	Sequence 1, Appli
c	35	13	52.0	2214	3	US-08-149-223A-1	Sequence 1, Appli
c	36	13	52.0	2381	2	US-08-484-993B-9	Sequence 9, Appli
c	37	13	52.0	2381	2	US-08-484-158B-9	Sequence 9, Appli
c	38	13	52.0	2381	2	US-08-484-596A-9	Sequence 9, Appli
c	39	13	52.0	2381	2	US-08-480-150A-9	Sequence 9, Appli
c	40	13	52.0	2381	3	US-08-458-731-9	Sequence 9, Appli
c	41	13	52.0	2381	3	US-08-149-223A-9	Sequence 9, Appli
c	42	13	52.0	3083	1	US-07-718-575-5	Sequence 5, Appli
c	43	13	52.0	3083	1	US-08-481-206-5	Sequence 5, Appli
c	44	13	52.0	3083	2	US-08-486-269A-5	Sequence 5, Appli
c	45	13	52.0	4500	2	US-08-743-637B-35	Sequence 35, Appl
c	46	13	52.0	4500	3	US-08-526-840B-35	Sequence 35, Appl
	47	13	52.0	5590	4	US-09-050-159-129	Sequence 129, App
c	48	13	52.0	72928	3	US-09-009-913-1	Sequence 1, Appli
c	49	13	52.0	80246	4	US-09-078-294-4	Sequence 4, Appli
c	50	13	52.0	80595	4	US-09-078-294-3	Sequence 3, Appli
	51	12	48.0	17	1	US-08-373-124A-284	Sequence 284, App
	52	12	48.0	17	1	US-08-373-124A-286	Sequence 286, App
	53	12	48.0	17	1	US-08-373-124A-2333	Sequence 2333, Ap
	54	12	48.0	17	1	US-08-373-124A-2335	Sequence 2335, Ap
	55	12	48.0	17	1	US-08-435-628-284	Sequence 284, App
	56	12	48.0	17	1	US-08-435-628-286	Sequence 286, App
	57	12	48.0	17	1	US-08-435-628-2333	Sequence 2333, Ap
	58	12	48.0	17	1	US-08-435-628-2335	Sequence 2335, Ap
c	59	12	48.0	24	3	US-09-014-065-15	Sequence 15, Appl
	60	12	48.0	27	1	US-08-373-124A-16	Sequence 16, Appl
	61	12	48.0	27	1	US-08-435-628-16	Sequence 16, Appl
	62	12	48.0	99	1	US-07-795-859B-22	Sequence 22, Appl
	63	12	48.0	99	1	US-08-457-616-22	Sequence 22, Appl
	64	12	48.0	184	2	US-08-485-657A-8	Sequence 8, Appli
	65	12	48.0	184	5	PCT-US95-02303-8	Sequence 8, Appli
	66	12	48.0	255	5	PCT-US93-03077-7	Sequence 7, Appli
c	67	12	48.0	279	1	US-08-047-041A-17	Sequence 17, Appl
c	68	12	48.0	308	4	US-09-232-522-3	Sequence 3, Appli
c	69	12	48.0	311	1	US-08-047-041A-4	Sequence 4, Appli
c	70	12	48.0	311	2	US-08-795-006A-24	Sequence 24, Appl
c	71	12	48.0	311	4	US-09-184-073-24	Sequence 24, Appl
	72	12	48.0	318	1	US-08-248-474-51	Sequence 51, Appl
	73	12	48.0	318	3	US-08-756-849-51	Sequence 51, Appl
c	74	12	48.0	354	3	US-09-065-059-12	Sequence 12, Appl
c	75	12	48.0	357	3	US-08-767-128-17	Sequence 17, Appl
c	76	12	48.0	411	1	US-08-137-117D-30	Sequence 30, Appl
c	77	12	48.0	411	1	US-08-436-717-30	Sequence 30, Appl
c	78	12	48.0	420	4	US-09-214-095D-109	Sequence 109, App
c	79	12	48.0	426	3	US-08-554-840-28	Sequence 28, Appl
c	80	12	48.0	427	2	US-08-484-956-111	Sequence 111, App
c	81	12	48.0	427	2	US-08-757-653-111	Sequence 111, App
c	82	12	48.0	498	2	US-08-484-956-113	Sequence 113, App
c	83	12	48.0	498	2	US-08-757-653-113	Sequence 113, App
	84	12	48.0	500	4	US-08-642-274D-18	Sequence 18, Appl
	85	12	48.0	500	4	US-08-952-014C-18	Sequence 18, Appl
	86	12	48.0	519	2	US-08-581-528A-5	Sequence 5, Appli
	87	12	48.0	519	5	PCT-US94-07799-5	Sequence 5, Appli
c	88	12	48.0	600	1	US-08-540-077-1	Sequence 1, Appli
c	89	12	48.0	601	2	US-08-484-956-97	Sequence 97, Appl
	90	12	48.0	601	2	US-08-484-956-98	Sequence 98, Appl
c	91	12	48.0	601	2	US-08-484-956-99	Sequence 99, Appl
	92	12	48.0	601	2	US-08-484-956-100	Sequence 100, App
c	93	12	48.0	601	2	US-08-484-956-101	Sequence 101, App
	94	12	48.0	601	2	US-08-484-956-102	Sequence 102, App
c	95	12	48.0	601	2	US-08-484-956-107	Sequence 107, App
	96	12	48.0	601	2	US-08-484-956-108	Sequence 108, App
c	97	12	48.0	601	2	US-08-484-956-109	Sequence 109, App
	98	12	48.0	601	2	US-08-484-956-110	Sequence 110, App
c	99	12	48.0	601	2	US-08-757-653-97	Sequence 97, Appl
	100	12	48.0	601	2	US-08-757-653-98	Sequence 98, Appl

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db      24 ATTGTCGCAAGTCACAGAATTATTT 48


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db      24 ATTGTCGCAAGTCACAGAATTATTT 48


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db     140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcacagaattattt 25
           |||||||||||||||||||||||||
Db     140 ATTGTCGCAAGTCACAGAATTATTT 164


RESULT   5
US-07-944-259-2/c
; Sequence 2, Application US/07944259
; Patent No. 5885792
;  GENERAL INFORMATION:
;    APPLICANT:  Ifuku, Ohji
;    APPLICANT:  Haze, Shinitiro
;    APPLICANT:  Kishimoto, Jiro
;    APPLICANT:  Nakahama, Kazuo
;    TITLE OF INVENTION:  BIOTIN OPERON
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wegner, Cantor, Mueller & Player
;      STREET:  1233 20th Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20036-8218
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/944,259
;      FILING DATE:  19920914
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cantor, Herbert I.
;      REGISTRATION NUMBER:  24,392
;      REFERENCE/DOCKET NUMBER:  P-450-23557
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-887-0400
;      TELEFAX:  202-835-0605
;      TELEX:  440706
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
US-07-944-259-2


  Query Match             68.0%;  Score 17;  DB 2;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 0.45;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gtcgcaagtcacagaat 20
           |||||||||||||||||
Db      17 GTCGCAAGTCACAGAAT 1


RESULT   6
US-09-166-203-48
; Sequence 48, Application US/09166203A
; Patent No. 5968826
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Condon, Tom P.
;  APPLICANT: Cowsert, Lex M.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
;  FILE REFERENCE: ISPH-0323
;  CURRENT APPLICATION NUMBER: US/09/166,203A
;  CURRENT FILING DATE: 1998-10-05
;  NUMBER OF SEQ ID NOS: 60
; SEQ ID NO 48
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1193)..(1387)
;   FEATURE:
```

```
;   NAME/KEY: CDS
;   LOCATION: (1709)..(1771)
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: L20788 Genbank
;   DATABASE ENTRY DATE: 1996-04-18
US-09-166-203-48


  Query Match             60.0%;  Score 15;  DB 2;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaatt 21
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT   7
US-09-377-309-48
; Sequence 48, Application US/09377309B
; Patent No. 6258790
; GENERAL INFORMATION:
;  APPLICANT: Bennett, C. Frank
;  APPLICANT: Condon, Tom P.
;  APPLICANT: Cowsert, Lex M.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF INTEGRIN 4 EXPRESSION
;  FILE REFERENCE: ISPH-0390
;  CURRENT APPLICATION NUMBER: US/09/377,309B
;  CURRENT FILING DATE: 1999-08-19
;  EARLIER APPLICATION NUMBER: 09/166,203
;  EARLIER FILING DATE: 1998-10-05
;  NUMBER OF SEQ ID NOS: 99
; SEQ ID NO 48
;   LENGTH: 1771
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1193)..(1387)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1709)..(1771)
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: L20788 Genbank
;   DATABASE ENTRY DATE: 1996-04-18
US-09-377-309-48


  Query Match             60.0%;  Score 15;  DB 4;  Length 1771;
  Best Local Similarity   100.0%;  Pred. No. 4;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaatt 21
           |||||||||||||||
Db     312 gcaagtcacagaatt 326


RESULT   8
5498529-7/c
;Patent No. 5498529
;     APPLICANT: BERKA, THOMAS R.;FORNWALD, JAMES A.;CORNIAK,
;JOSELINA G.;ROSENBERG, MARTIN;STRICKLER, JAMES E.;TAYLOR, DEAN P.
;     TITLE OF INVENTION: PROTEIN PROTEASE INHIBITORS FROM
;STREPTOMYCES
;     NUMBER OF SEQUENCES: 9
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/162,506
;       FILING DATE: 03-DEC-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 894,167
;       FILING DATE: 03-JUN-1992
;       APPLICATION NUMBER: 346,119
;       FILING DATE: 16-FEB-1989
;       APPLICATION NUMBER: 897,245
;       FILING DATE: 18-AUG-1986
;SEQ ID NO:7:
;       LENGTH: 1050
5498529-7


  Query Match             56.0%;  Score 14;  DB 6;  Length 1050;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatta 22
           ||||||||||||||
Db     251 AAGTCACAGAATTA 238


RESULT   9
US-08-372-652-14/c
; Sequence 14, Application US/08372652
; Patent No. 5932699
;  GENERAL INFORMATION:
;     APPLICANT:  Moore, David
;     APPLICANT:  Seol, Wongi
;     APPLICANT:  Choi, Hueng-Sik
;     TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;     TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street, Suite 3100
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/372,652
;       FILING DATE:  13-JAN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Paul T.
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  00786/246001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;  INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1677 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-372-652-14


  Query Match             56.0%;  Score 14;  DB 2;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgtcgcaagtcac 15
           ||||||||||||||
Db     630 TTGTCGCAAGTCAC 617


RESULT  10
PCT-US95-16311-14/c
; Sequence 14, Application PC/TUS9516311
;  GENERAL INFORMATION:
```

```
;     APPLICANT:  Moore, David
;     APPLICANT:  Seol, Wongi
;     APPLICANT:  Choi, Hueng-Sik
;     TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;     TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street, Suite 3100
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/16311
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/372,652
;       FILING DATE:  13-JAN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Paul T.
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  00786/246001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  14:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1677 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
PCT-US95-16311-14


  Query Match             56.0%;  Score 14;  DB 5;  Length 1677;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgtcgcaagtcac 15
           ||||||||||||||
Db     630 TTGTCGCAAGTCAC 617


RESULT  11
US-08-372-652-6/c
; Sequence 6, Application US/08372652
; Patent No. 5932699
;  GENERAL INFORMATION:
;     APPLICANT:  Moore, David
;     APPLICANT:  Seol, Wongi
;     APPLICANT:  Choi, Hueng-Sik
;     TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;     TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street, Suite 3100
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/372,652
;       FILING DATE:  13-JAN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Paul T.
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  00786/246001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1787 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
US-08-372-652-6


  Query Match             56.0%;  Score 14;  DB 2;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgtcgcaagtcac 15
           ||||||||||||||
Db     740 TTGTCGCAAGTCAC 727


RESULT  12
PCT-US95-16311-6/c
; Sequence 6, Application PC/TUS9516311
;  GENERAL INFORMATION:
;     APPLICANT:  Moore, David
;     APPLICANT:  Seol, Wongi
;     APPLICANT:  Choi, Hueng-Sik
;     TITLE OF INVENTION:  RETINOID X RECEPTOR-INTERACTING
;     TITLE OF INVENTION:  POLYPEPTIDES AND RELATED MOLECULES AND METHODS
;     NUMBER OF SEQUENCES:  17
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street, Suite 3100
;       CITY:  Boston
;       STATE:  MA
;       COUNTRY:  USA
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/16311
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/372,652
;       FILING DATE:  13-JAN-1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Paul T.
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  00786/246001
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/542-5070
;       TELEFAX:  617/542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1787 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  DNA
PCT-US95-16311-6


  Query Match             56.0%;  Score 14;  DB 5;  Length 1787;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db    740 TTGTCGCAAGTCAC 727


RESULT  13
US-08-765-081-4
; Sequence 4, Application US/08765081
; Patent No. 5798260
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium
;      OPERATING SYSTEM:  MS-Window 3.1
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,081
;      FILING DATE:  March 26, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997
;      REFERENCE/DOCKET NUMBER:  CHOR-1-10286
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0718 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  Corresponds to SEQ ID NO:1,
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
US-08-765-081-4


  Query Match             56.0%;  Score 14;  DB 1;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db     692 AGTCACAGAATTAT 705


RESULT  14
US-09-098-082-4
; Sequence 4, Application US/09098082
; Patent No. 6040427
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium II
;      OPERATING SYSTEM:  MS-Windows 95
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/098,082
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/765,081
;      FILING DATE:  March 26, 1997
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  CHOR-1-12402
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0735 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  Corresponds to SEQ ID NO:1,
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
US-09-098-082-4


  Query Match             56.0%;  Score 14;  DB 3;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db     692 AGTCACAGAATTAT 705


RESULT  15
PCT-US95-06994-4
; Sequence 4, Application PC/TUS9506994
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Children's Hospital & Medical Center
;    APPLICANT:  University of Washington
;    APPLICANT:  Washington State University Research Foundation
;    APPLICANT:  TARR, PHILLIP I
;    APPLICANT:  BILGE, SIMA S
;    APPLICANT:  BESSER, THOMAS E
;    APPLICANT:  VARY JR, JAMES C
;    TITLE OF INVENTION:  ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
;      STREET:  SUITE 2800, 1420 FIFTH AVENUE
;      CITY:  SEATTLE
;      STATE:  WASHINGTON
;      COUNTRY:  USA
;      ZIP:  WA 98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  07-JUN-95
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  24-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BRODERICK, THOMAS F
;      REGISTRATION NUMBER:  31,332
;      REFERENCE/DOCKET NUMBER:  CHOR-18591
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682 8100
;      TELEFAX:  (206) 224 0779
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;      DESCRIPTION:  corresponds to SEQ ID NO:1,
;      DESCRIPTION:  nucleotides 3036-5126
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157: H7
;      STRAIN:  86-24 NALR
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2088
PCT-US95-06994-4


  Query Match             56.0%;  Score 14;  DB 5;  Length 2091;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db     692 AGTCACAGAATTAT 705


RESULT  16
US-08-496-631-1/c
; Sequence 1, Application US/08496631
; Patent No. 5728548'
;  GENERAL INFORMATION:
;    APPLICANT:  Bowman, Michael
;    TITLE OF INVENTION:  STEROID RECEPTOR RR1
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/496,631
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5248
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2233 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  363..1778
US-08-496-631-1


  Query Match             56.0%;  Score 14;  DB 1;  Length 2233;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgtcgcaagtcac 15
          ||||||||||||||
Db   1055 TTGTCGCAAGTCAC 1042


RESULT  17
US-09-008-697A-7/c
; Sequence 7, Application US/09008697A
; Patent No. 6197504
;  GENERAL INFORMATION:
;    APPLICANT:  Chow, King Lau
;    TITLE OF INVENTION:  USES OF MAB-21
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/008,697A
;      FILING DATE:  January 19, 1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Chan, Albert Wai-Kit
;      REGISTRATION NUMBER:  36,479
```

```
;      REFERENCE/DOCKET NUMBER:  50752-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2585 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-008-697A-7


  Query Match             56.0%;  Score 14;  DB 4;  Length 2585;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatt 21
          ||||||||||||||
Db   1058 CAAGTCACAGAATT 1045


RESULT  18
US-08-765-081-1
; Sequence 1, Application US/08765081
; Patent No. 5798260
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium
;      OPERATING SYSTEM:  MS-Window 3.1
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,081
;      FILING DATE:  March 26, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997
;      REFERENCE/DOCKET NUMBER:  CHOR-1-10286
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0718 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157:H7
;      STRAIN:  86-24 NALR
;    IMMEDIATE SOURCE:
;      CLONE:  pEAR
US-08-765-081-1


  Query Match             56.0%;  Score 14;  DB 1;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db   3727 AGTCACAGAATTAT 3740


RESULT  19
US-09-098-082-1
; Sequence 1, Application US/09098082
; Patent No. 6040421
;  GENERAL INFORMATION:
;    APPLICANT:  Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
;    TITLE OF INVENTION:  Escherichia Coli 0157:H7 Epithelial Adhesin
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson and Kindness
;      STREET:  2800 Pacific First Center, 1420 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette-3.5 inch, 1.44Mb storage
;      COMPUTER:  IBM PC compatible/Pentium II
;      OPERATING SYSTEM:  MS-Windows 95
;      SOFTWARE:  Word for Windows-6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/098,082
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/765,081
;      FILING DATE:  March 26, 1997
;      APPLICATION NUMBER:  PCT/US95/06994
;      FILING DATE:  June 7, 1995
;      APPLICATION NUMBER:  US 08/265,714
;      FILING DATE:  June 24, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  CHOR-1-12402
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  1-206-682-8100; 1-206-224-0735 (direct)
;      TELEFAX:  1-206-224-0779
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli O157:H7
;      STRAIN:  86-24 NALR
;    IMMEDIATE SOURCE:
;      CLONE:  pEAR
US-09-098-082-1


  Query Match             56.0%;  Score 14;  DB 3;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db   3727 AGTCACAGAATTAT 3740
```

```
RESULT  20
PCT-US95-06994-1
; Sequence 1, Application PC/TUS9506994
;  GENERAL INFORMATION:
;     APPLICANT:  Children's Hospital & Medical Center
;     APPLICANT:  University of Washington
;     APPLICANT:  Washington State University Research Foundation
;     APPLICANT:  TARR, PHILLIP I
;     APPLICANT:  BILGE, SIMA S
;     APPLICANT:  BESSER, THOMAS E
;     APPLICANT:  VARY JR, JAMES C
;     TITLE OF INVENTION:  ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
;     NUMBER OF SEQUENCES:  14
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
;       STREET:  SUITE 2800, 1420 FIFTH AVENUE
;       CITY:  SEATTLE
;       STATE:  WASHINGTON
;       COUNTRY:  USA
;       ZIP:  WA 98101
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/06994
;       FILING DATE:  07-JUN-95
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/265,714
;       FILING DATE:  24-JUN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  BRODERICK, THOMAS F
;       REGISTRATION NUMBER:  31,332
;       REFERENCE/DOCKET NUMBER:  CHOR-18591
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (206) 682 8100
;       TELEFAX:  (206) 224 0779
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8041 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Escherichia coli O157:H7
;       STRAIN:  86-24 NALR
;     IMMEDIATE SOURCE:
;       CLONE:  pEAR
PCT-US95-06994-1


  Query Match             56.0%;  Score 14;  DB 5;  Length 8041;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattat 23
           ||||||||||||||
Db    3727 AGTCACAGAATTAT 3740


RESULT  21
US-09-385-982-372/c
; Sequence 372, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;  APPLICANT: ENDEGE, WILSON O., ET AL.
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS:  II
;  FILE REFERENCE: CCDNA-260XX
;  CURRENT APPLICATION NUMBER: US/09/385,982
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: 09/328,111
;  EARLIER FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: 60/117,393
;  EARLIER FILING DATE: 1999-01-27
;  EARLIER APPLICATION NUMBER: 60/098,639
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 544
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 372
;   LENGTH: 607
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(607)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-372


  Query Match             52.0%;  Score 13;  DB 4;  Length 607;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattatt 24
           |||||||||||||
Db    189 TCACAGAATTATT 177


RESULT  22
US-09-385-982-250
; Sequence 250, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;  APPLICANT: ENDEGE, WILSON O., ET AL.
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS:  II
;  FILE REFERENCE: CCDNA-260XX
;  CURRENT APPLICATION NUMBER: US/09/385,982
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: 09/328,111
;  EARLIER FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: 60/117,393
;  EARLIER FILING DATE: 1999-01-27
;  EARLIER APPLICATION NUMBER: 60/098,639
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 544
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 250
;   LENGTH: 626
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(626)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-250


  Query Match             52.0%;  Score 13;  DB 4;  Length 626;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattatt 24
           |||||||||||||
Db    177 tcacagaattatt 189


RESULT  23
US-09-385-982-253
```

```
; Sequence 253, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;  APPLICANT: ENDEGE, WILSON O., ET AL.
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS:  II
;  FILE REFERENCE: CCDNA-260XX
;  CURRENT APPLICATION NUMBER: US/09/385,982
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: 09/328,111
;  EARLIER FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: 60/117,393
;  EARLIER FILING DATE: 1999-01-27
;  EARLIER APPLICATION NUMBER: 60/098,639
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 544
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 253
;   LENGTH: 634
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(634)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-253


  Query Match            52.0%;  Score 13;  DB 4;  Length 634;
  Best Local Similarity  100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     12 tcacagaattatt 24
          |||||||||||||
Db    196 tcacagaattatt 208


RESULT  24
US-09-385-982-329
; Sequence 329, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;  APPLICANT: ENDEGE, WILSON O., ET AL.
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS:  II
;  FILE REFERENCE: CCDNA-260XX
;  CURRENT APPLICATION NUMBER: US/09/385,982
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: 09/328,111
;  EARLIER FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: 60/117,393
;  EARLIER FILING DATE: 1999-01-27
;  EARLIER APPLICATION NUMBER: 60/098,639
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 544
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 329
;   LENGTH: 651
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(651)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-329


  Query Match            52.0%;  Score 13;  DB 4;  Length 651;
  Best Local Similarity  100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     12 tcacagaattatt 24
          |||||||||||||
Db    169 tcacagaattatt 181


RESULT  25
US-08-875-847B-3/c
; Sequence 3, Application US/08875847B
; Patent No. 6255105
;  GENERAL INFORMATION:
;    APPLICANT:  The Government of the United
;    APPLICANT:  States of America as represented by the
;    APPLICANT:  Secretary, Department of Health and Human
;    APPLICANT:  Services; Callahan, Robert; Marchetti,
;    APPLICANT:  Antonio; Buttitta, Fiamma; Smith, Gilbert H.
;    TITLE OF INVENTION:  Nucleotide And Deduced
;    TITLE OF INVENTION:  Amino Acid Sequences Of A New Tumor Gene,
;    TITLE OF INVENTION:  Int6, And the Use Of Reagents Derived From
;    TITLE OF INVENTION:  These Sequences In Diagnostic Assays,
;    TITLE OF INVENTION:  Vaccines, Immunotherapy And Gene Therapy
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN, L.L.P.
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MS WORD 97
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/875,847B
;      FILING DATE:  09-FEB-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/385,998
;      FILING DATE:  09-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William S. Feiler
;      REGISTRATION NUMBER:  26,728
;      REFERENCE/DOCKET NUMBER:  2026-4179PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;      TELEX:  421792
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1500 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-875-847B-3


  Query Match            52.0%;  Score 13;  DB 4;  Length 1500;
  Best Local Similarity  100.0%;  Pred. No. 49;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     12 tcacagaattatt 24
          |||||||||||||
Db    624 TCACAGAATTATT 612


RESULT  26
US-08-875-847B-1/c
; Sequence 1, Application US/08875847B
; Patent No. 6255105
;  GENERAL INFORMATION:
;    APPLICANT:  The Government of the United
;    APPLICANT:  States of America as represented by the
;    APPLICANT:  Secretary, Department of Health and Human
```

```
;     APPLICANT:  Services; Callahan, Robert; Marchetti,
;     APPLICANT:  Antonio; Buttitta, Fiamma; Smith, Gilbert H.
;     TITLE OF INVENTION:  Nucleotide And Deduced
;     TITLE OF INVENTION:  Amino Acid Sequences Of A New Tumor Gene,
;     TITLE OF INVENTION:  Int6, And the Use Of Reagents Derived From
;     TITLE OF INVENTION:  These Sequences In Diagnostic Assays,
;     TITLE OF INVENTION:  Vaccines, Immunotherapy And Gene Therapy
;     NUMBER OF SEQUENCES:  32
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORGAN & FINNEGAN, L.L.P.
;       STREET:  345 PARK AVENUE
;       CITY:  NEW YORK
;       STATE:  NEW YORK
;       COUNTRY:  USA
;       ZIP:  10154
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  FLOPPY DISK
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  MS WORD 97
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/875,847B
;       FILING DATE:  09-FEB-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/385,998
;       FILING DATE:  09-FEB-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  William S. Feiler
;       REGISTRATION NUMBER:  26,728
;       REFERENCE/DOCKET NUMBER:  2026-4179PCT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 758-4800
;       TELEFAX:  (212) 751-6849
;       TELEX:  421792
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1505 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-875-847B-1


  Query Match             52.0%;  Score 13;  DB 4;  Length 1505;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 tcacagaattatt 24
           |||||||||||||
Db     630 TCACAGAATTATT 618


RESULT  27
US-08-651-572-1/c
; Sequence 1, Application US/08651572
; Patent No. 5789228
;  GENERAL INFORMATION:
;     APPLICANT:  Lam, D. et al.
;     TITLE OF INVENTION:  Endoglucanases
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;       ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;       STREET:  6 BECKER FARM ROAD
;       CITY:  ROSELAND
;       STATE:  NEW JERSEY
;       COUNTRY:  USA
;       ZIP:  07068
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 INCH DISKETTE
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  WORD PERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/651,572
;       FILING DATE:  Herewith
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Charles J. Herron
;       REGISTRATION NUMBER:  28,019
;       REFERENCE/DOCKET NUMBER:  331400-48
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-994-1700
;       TELEFAX:  201-994-1744
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1662 NUCLEOTIDES
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  SINGLE
;       TOPOLOGY:  LINEAR
;     MOLECULE TYPE:  DNA
US-08-651-572-1


  Query Match             52.0%;  Score 13;  DB 1;  Length 1662;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatt 21
           |||||||||||||
Db     974 AAGTCACAGAATT 962


RESULT  28
US-09-066-544-1/c
; Sequence 1, Application US/09066544
; Patent No. 6001984
;  GENERAL INFORMATION:
;     APPLICANT:  Lam, D. et al.
;     TITLE OF INVENTION:  Endoglucanases
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;       ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;       STREET:  6 BECKER FARM ROAD
;       CITY:  ROSELAND
;       STATE:  NEW JERSEY
;       COUNTRY:  USA
;       ZIP:  07068
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5 INCH DISKETTE
;       COMPUTER:  IBM PS/2
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  WORD PERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/066,544
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/651,572
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Charles J. Herron
;       REGISTRATION NUMBER:  28,019
;       REFERENCE/DOCKET NUMBER:  331400-48
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  201-994-1700
;       TELEFAX:  201-994-1744
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1662 NUCLEOTIDES
```

```
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-09-066-544-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 1662;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatt 21
           |||||||||||||
Db     974 AAGTCACAGAATT 962


RESULT  29
US-08-951-086-1/c
; Sequence 1, Application US/08951086
; Patent No. 6074867
;  GENERAL INFORMATION:
;    APPLICANT:  Lam, D. et al.
;    TITLE OF INVENTION:  Endoglucanases
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN, GILFILLAN,
;      ADDRESSEE:  CECCHI, STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NEW JERSEY
;      COUNTRY:  USA
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 INCH DISKETTE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  WORD PERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/951,086
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/651,572
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles J. Herron
;      REGISTRATION NUMBER:  28,019
;      REFERENCE/DOCKET NUMBER:  331400-48
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1662 NUCLEOTIDES
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  SINGLE
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:  DNA
US-08-951-086-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 1662;
  Best Local Similarity   100.0%;  Pred. No. 49;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatt 21
           |||||||||||||
Db     974 AAGTCACAGAATT 962


RESULT  30
US-08-484-993B-1/c
; Sequence 1, Application US/08484993B
; Patent No. 5837497
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,993B
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-484-993B-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886
```

```
RESULT  31
US-08-484-158B-1/c
; Sequence 1, Application US/08484158B
; Patent No. 5976545
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Pharmaceutical Compositions for
;    TITLE OF INVENTION:  Immunocontraception
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,158B
;      FILING DATE:  07-JUNE-95
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  09-NOV-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  32794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-484-158B-1


  Query Match           52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  32
US-08-484-596A-1/c
; Sequence 1, Application US/08484596A
; Patent No. 5981228
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,596A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  11-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-484-596A-1
```

```
  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  33
US-08-480-150A-1/c
; Sequence 1, Application US/08480150A
; Patent No. 5989550
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,150A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/149,223
;      FILING DATE:  09-NOV-1993
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-480-150A-1


  Query Match             52.0%;  Score 13;  DB 2;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  34
US-08-458-731-1/c
; Sequence 1, Application US/08458731
; Patent No. 6001599
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,731
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
```

```
;       LOCATION:  12..119
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  120..2153
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  12..2153
US-08-458-731-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  35
US-08-149-223A-1/c
; Sequence 1, Application US/08149223A
; Patent No. 6027727
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/149,223A
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2214 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Sus scrofa
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  12..119
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  120..2153
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  12..2153
US-08-149-223A-1


  Query Match             52.0%;  Score 13;  DB 3;  Length 2214;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    1898 GCAAGTCACAGAA 1886


RESULT  36
US-08-484-993B-9/c
; Sequence 9, Application US/08484993B
; Patent No. 5837497
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,993B
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
```

```
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-484-993B-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  37
US-08-484-158B-9/c
; Sequence 9, Application US/08484158B
; Patent No. 5976545
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Pharmaceutical Compositions for
;    TITLE OF INVENTION:  Immunocontraception
;    NUMBER OF SEQUENCES:  61
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,158B
;      FILING DATE:  07-JUNE-95
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  09-NOV-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-93
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  32794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-484-158B-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  38
US-08-484-596A-9/c
; Sequence 9, Application US/08484596A
; Patent No. 5981228
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,596A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/149,223
;      FILING DATE:  11-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
```

```
;       HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  206..2353
US-08-484-596A-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  39
US-08-480-150A-9/c
; Sequence 9, Application US/08480150A
; Patent No. 5989550
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,150A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/149,223
;      FILING DATE:  09-NOV-1993
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;       TISSUE TYPE:  Ovary
;       CELL TYPE:  Oocyte
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  206..2353
US-08-480-150A-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  40
US-08-458-731-9/c
; Sequence 9, Application US/08458731
; Patent No. 6001599
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,731
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
```

```
;       NAME/KEY:  CDS
;       LOCATION:  206..2353
US-08-458-731-9


  Query Match             52.0%;  Score 13;  DB 3;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  41
US-08-149-223A-9/c
; Sequence 9, Application US/08149223A
; Patent No. 6027727
;  GENERAL INFORMATION:
;    APPLICANT:  Harris Ph.D., Jeffrey D.
;    APPLICANT:  Hsu, Kuang T.
;    APPLICANT:  Podolski, Joseph S.
;    TITLE OF INVENTION:  Materials and Methods for Immunocontraception
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/149,223A
;      FILING DATE:  09-NOV-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/012,990
;      FILING DATE:  29-JAN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/973,341
;      FILING DATE:  09-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  31745
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6653
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2381 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Canis familiaris
;      DEVELOPMENTAL STAGE:  Juvenile
;      HAPLOTYPE:  Diploidy
;      TISSUE TYPE:  Ovary
;      CELL TYPE:  Oocyte
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  206..2353
US-08-149-223A-9


  Query Match             52.0%;  Score 13;  DB 3;  Length 2381;
  Best Local Similarity   100.0%;  Pred. No. 48;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaa 19
           |||||||||||||
Db    2098 GCAAGTCACAGAA 2086


RESULT  42
US-07-718-575-5/c
; Sequence 5, Application US/07718575
; Patent No. 5202257
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann Ph.D., Stephen F.
;    APPLICANT:  Boulter Ph.D., James R.
;    APPLICANT:  Hollmann Ph.D., Michael NMN
;    APPLICANT:  Bettler Ph.D., Bernhard NMN
;    APPLICANT:  Jensen Ph.D., Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 So. Flower St., Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071-2921
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/718,575
;      FILING DATE:  19910813
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter Ph.D., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  P31 8962
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619)535-8949
;      TELEX:  9103330318
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3083 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  167..2833
US-07-718-575-5


  Query Match             52.0%;  Score 13;  DB 1;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtc 13
           |||||||||||||
Db    3033 ATTGTCGCAAGTC 3021


RESULT  43
```

```
US-08-481-206-5/c
; Sequence 5, Application US/08481206
; Patent No. 5739291
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann Ph.D., Stephen F.
;    APPLICANT:  Boulter Ph.D., James R.
;    APPLICANT:  Hollmann Ph.D., Michael NMN
;    APPLICANT:  Bettler Ph.D., Bernhard NMN
;    APPLICANT:  Jensen Ph.D., Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS AND
;    TITLE OF INVENTION:  METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 So. Flower St., Suite 2000
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  90071-2921
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/481,206
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/013,767
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter Ph.D., Stephen E.
;      REGISTRATION NUMBER:  31192
;      REFERENCE/DOCKET NUMBER:  P31 8962
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619)535-8949
;      TELEX:  9103330318
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3083 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  167..2833
US-08-481-206-5


  Query Match             52.0%;  Score 13;  DB 1;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtc 13
           |||||||||||||
Db    3033 ATTGTCGCAAGTC 3021


RESULT  44
US-08-486-269A-5/c
; Sequence 5, Application US/08486269A
; Patent No. 5945509
;  GENERAL INFORMATION:
;    APPLICANT:  Heinemann, Stephen F.
;    APPLICANT:  Boulter, James R.
;    APPLICANT:  Hollmann, Michael
;    APPLICANT:  Bettler, Bernhard
;    APPLICANT:  Jensen, Jan E.
;    TITLE OF INVENTION:  GLUTAMATE RECEPTOR COMPOSITIONS
;    TITLE OF INVENTION:  AND METHODS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gray Cary Ware & Freidenrich LLP
;      STREET:  4365 Executive Drive, Suite 1600
;      CITY:  San Diego
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92121
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows DEMONSTRATION Version 2.0D
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/486,269A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/013,767
;      FILING DATE:  04-FEB-1993
;      APPLICATION NUMBER:  07/718,575
;      FILING DATE:  21-JUN-1991
;      APPLICATION NUMBER:  PCT/US90/06153
;      FILING DATE:  25-OCT-1990
;      APPLICATION NUMBER:  07/428,116
;      FILING DATE:  27-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9986
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-677-1409
;      TELEFAX:  619-677-1465
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3083 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  GluR3
;    FEATURE:
;      NAME/KEY:  Coding Sequence
;      LOCATION:  167...2830
;      OTHER INFORMATION:
US-08-486-269A-5


  Query Match             52.0%;  Score 13;  DB 2;  Length 3083;
  Best Local Similarity   100.0%;  Pred. No. 47;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtc 13
           |||||||||||||
Db    3033 ATTGTCGCAAGTC 3021


RESULT  45
US-08-743-637B-35/c
; Sequence 35, Application US/08743637B
; Patent No. 5994066
;  GENERAL INFORMATION:
;    APPLICANT:  BERGERON, Michel G.
;    APPLICANT:  PICARD, Francois J.
;    APPLICANT:  OUELLETTE, Marc
;    APPLICANT:  ROY, Paul H.
;    TITLE OF INVENTION:  SPECIES-SPECIFIC AND UNIVERSAL DNA
;    TITLE OF INVENTION:  PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
;    TITLE OF INVENTION:  IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
```

```
;    TITLE OF INVENTION:  ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
;    NUMBER OF SEQUENCES:  273
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  QUARLES & BRADY
;      STREET:  411 EAST WISCONSIN AVENUE
;      CITY:  MILWAUKEE
;      STATE:  WISCONSIN
;      COUNTRY:  USA
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/743,637B
;      FILING DATE:  04-NOV-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/526,840
;      FILING DATE:  11-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BAKER, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  850586.90012
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5000
;      TELEFAX:  (414)277-5591
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4500 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus pneumoniae
US-08-743-637B-35


  Query Match            52.0%;  Score 13;  DB 2;  Length 4500;
  Best Local Similarity  100.0%;  Pred. No. 46;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaatt 21
           |||||||||||||
Db    1911 AAGTCACAGAATT 1899



Search completed: December 26, 2001, 12:52:31
Job time: 6783 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:09:53 ; Search time 4619.78 Seconds
(without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196F-2
Perfect score: 25
Sequence: 1 attgtcgcaagtcacagaattattt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 9

Total number of hits satisfying chosen parameters: 637897

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	18	72.0	357	11	BE800706	BE800706 sq97c06.y
	2	18	72.0	388	10	AW781150	AW781150 sl89g06.y
c	3	18	72.0	751	10	BE658717	BE658717 GM700007A
	4	17	68.0	424	13	AQ227679	AQ227679 HS_2012_B
c	5	17	68.0	944	10	AI529608	AI529608 ui80b10.y
c	6	16	64.0	263	10	BB141316	BB141316 BB141316
c	7	16	64.0	270	10	BE181705	BE181705 QV1-HT063
	8	16	64.0	299	10	BB450863	BB450863 BB450863
	9	16	64.0	353	10	AA648449	AA648449 ns22f10.s
	10	16	64.0	405	11	BF726110	BF726110 by01c01.x
	11	16	64.0	429	11	BF232119	BF232119 dd99d06.x
	12	16	64.0	443	10	AI149091	AI149091 qc85f02.x
	13	16	64.0	454	13	AQ086771	AQ086771 HS_2175_B
	14	16	64.0	457	10	AI361386	AI361386 qy47f01.x
	15	16	64.0	467	11	BG360303	BG360303 dab93h11.
	16	16	64.0	471	10	AI635259	AI635259 tz79h07.x
	17	16	64.0	503	10	AI634722	AI634722 tz09e04.x
	18	16	64.0	503	11	BG347513	BG347513 dab87h12.
	19	16	64.0	512	10	AW418528	AW418528 xr32c07.x
	20	16	64.0	527	13	AQ472309	AQ472309 CITBI-E1-
c	21	16	64.0	649	11	BF322232	BF322232 maa34h05.
c	22	16	64.0	652	13	AQ541562	AQ541562 RPCI-11-3
c	23	16	64.0	657	11	BF726111	BF726111 by01c01.y
	24	16	64.0	702	13	AZ382380	AZ382380 1M0139H04
c	25	16	64.0	753	11	BG496183	BG496183 602538125
c	26	16	64.0	971	13	AQ900023	AQ900023 HS_3180_A
c	27	16	64.0	1318	11	BG871816	BG871816 602792574
	28	15	60.0	172	11	BG977588	BG977588 PM4-CI008
c	29	15	60.0	214	10	AW871559	AW871559 ra79d04.y
	30	15	60.0	233	10	BE763893	BE763893 RC4-NT005
	31	15	60.0	253	10	BE763892	BE763892 RC4-NT005
	32	15	60.0	256	10	AA317110	AA317110 EST19001
c	33	15	60.0	261	11	Z45844	Z45844 HSCZVE041 n
	34	15	60.0	263	11	BF745516	BF745516 CM2-BT082
c	35	15	60.0	266	10	AA115083	AA115083 zl06c10.r
c	36	15	60.0	282	11	BF800965	BF800965 QV4-CI010
	37	15	60.0	286	10	AA344205	AA344205 EST50086
c	38	15	60.0	292	13	B53227	B53227 CIT-HSP-200
c	39	15	60.0	317	11	BF743951	BF743951 CM2-BT082
c	40	15	60.0	324	13	BH121038	BH121038 RPCI-24-2
c	41	15	60.0	327	11	T30901	T30901 EST24536 Hu
c	42	15	60.0	335	11	BF743964	BF743964 CM2-BT082
c	43	15	60.0	346	11	BF743954	BF743954 CM2-BT082
	44	15	60.0	370	11	T07051	T07051 EST04940 Fe
c	45	15	60.0	381	11	H53286	H53286 yq85f09.r1
c	46	15	60.0	395	10	AA229851	AA229851 nc51e03.r
	47	15	60.0	402	10	AI694429	AI694429 wd83f06.x
	48	15	60.0	404	10	AI203847	AI203847 qf75h03.x
c	49	15	60.0	405	13	B95998	B95998 F21F2TFB IG
	50	15	60.0	416	10	AI942464	AI942464 wp80c05.x
	51	15	60.0	420	10	AA172188	AA172188 zp29f06.s
	52	15	60.0	426	11	BI003388	BI003388 PM0-HN007
c	53	15	60.0	427	13	AQ142435	AQ142435 HS_3064_A
c	54	15	60.0	429	10	AW870882	AW870882 ra51h09.y
	55	15	60.0	455	10	AA973493	AA973493 oo45e05.s
c	56	15	60.0	460	10	AV679467	AV679467 AV679467
	57	15	60.0	468	13	AQ116660	AQ116660 HS_3087_B
	58	15	60.0	473	10	AW882943	AW882943 QV2-OT006
	59	15	60.0	485	11	BG020076	BG020076 dc47c07.x
c	60	15	60.0	489	11	BF775935	BF775935 286217 MA
	61	15	60.0	498	13	AZ398330	AZ398330 1M0163N15
	62	15	60.0	502	10	AA846453	AA846453 aj56b03.s
	63	15	60.0	502	10	AI672463	AI672463 wa03e05.x
	64	15	60.0	503	10	AA824319	AA824319 aj30c08.s
c	65	15	60.0	505	13	AQ012645	AQ012645 CIT-HSP-2
c	66	15	60.0	511	13	AQ021225	AQ021225 CIT-HSP-2
	67	15	60.0	512	13	AQ801291	AQ801291 HS_2088_A
	68	15	60.0	520	13	AQ602419	AQ602419 HS_2128_A
	69	15	60.0	531	10	BE043069	BE043069 ho32e02.x
c	70	15	60.0	531	13	AQ806020	AQ806020 HS_3237_A
c	71	15	60.0	531	13	AZ908340	AZ908340 RPCI-24-2
	72	15	60.0	536	13	AZ027883	AZ027883 RPCI-23-3
c	73	15	60.0	538	10	AV677666	AV677666 AV677666
	74	15	60.0	548	13	AZ281398	AZ281398 RPCI-23-1
c	75	15	60.0	552	10	AL375298	AL375298 MtBB12F10
	76	15	60.0	554	10	AI268404	AI268404 qm05e10.x
c	77	15	60.0	562	13	AQ483592	AQ483592 RPCI-11-2
c	78	15	60.0	563	13	AQ430099	AQ430099 HS_5074_A
	79	15	60.0	585	10	AV674611	AV674611 AV674611
c	80	15	60.0	587	13	AZ554645	AZ554645 RPCI-23-1
c	81	15	60.0	596	11	N48234	N48234 yv22g01.r1
c	82	15	60.0	601	10	AW828925	AW828925 ra69g07.y
	83	15	60.0	617	10	AW316985	AW316985 xx10c10.x
	84	15	60.0	626	10	AV657973	AV657973 AV657973
c	85	15	60.0	658	13	AZ816912	AZ816912 2M0085C19

c 86	15	60.0	665	13	AZ459164	AZ459164 1M0263K18
87	15	60.0	691	13	AZ963904	AZ963904 2M0233D23
88	15	60.0	782	13	BH028572	BH028572 RPCI-24-3
c 89	15	60.0	849	10	BE548526	BE548526 601075921
c 90	15	60.0	874	11	BF239747	BF239747 601906343
91	15	60.0	912	10	AA626775	AA626775 ad09e07.s
92	15	60.0	1239	10	BE735021	BE735021 601569755
c 93	14	56.0	114	11	BF987752	BF987752 CM0-GN016
94	14	56.0	116	10	AA022548	AA022548 ze70f10.r
c 95	14	56.0	116	11	BF991166	BF991166 CM0-GN016
96	14	56.0	132	13	AZ073992	AZ073992 RPCI-23-4
97	14	56.0	142	10	BB184084	BB184084 BB184084
98	14	56.0	142	10	BB266371	BB266371 BB266371
99	14	56.0	149	10	BB300077	BB300077 BB300077
c 100	14	56.0	150	10	AW817267	AW817267 QV0-ST024

ALIGNMENTS

RESULT 1
BE800706
LOCUS BE800706 357 bp mRNA EST 20-SEP-2000
DEFINITION sq97c06.y1 Gm-c1049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1049-1091 5' similar to TR:O82150 O82150 CHLOROPLAST FTSH PROTEASE. ;, mRNA sequence.
ACCESSION BE800706
VERSION BE800706.1 GI:10231818
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 357)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Putative full length read
vector to vector length is 364
High quality sequence stop: 340.
FEATURES Location/Qualifiers
source 1. .357
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1049-1091"
/clone_lib="Gm-c1049"
/tissue_type="whole seedlings of greenhouse grown plants"
/dev_stage="3 week old"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The Clark NIL was constructed and seed was provided by Dr. J. Specht, University of Nebraska (Shoemaker and Specht. 1995). The cDNA library was constructed from mRNA isolated from whole seedlings of 3 week old greenhouse grown plants. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site and a 3' anchor. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). The library was constructed in cooperation with Dr. Paul Keim's laboratory at Northern Arizona University."
BASE COUNT 169 a 53 c 47 g 88 t
ORIGIN

Query Match 72.0%; Score 18; DB 11; Length 357;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      6 cgcaagtcacagaattat 23
          ||||||||||||||||||
Db    138 CGCAAGTCACAGAATTAT 155
```

RESULT 2
AW781150
LOCUS AW781150 388 bp mRNA EST 12-MAY-2000
DEFINITION sl89g06.y1 Gm-c1037 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1037-1235 5' similar to TR:O82150 O82150 CHLOROPLAST FTSH PROTEASE. ;, mRNA sequence.
ACCESSION AW781150
VERSION AW781150.1 GI:7795753
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 388)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 375.
FEATURES Location/Qualifiers
source 1. .388
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1037-1235"
/clone_lib="Gm-c1037"
/tissue_type="fully expanded leaves of greenhouse grown plants"
/dev_stage="2 week old"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This cDNA library was constructed from mRNA isolated from fully expanded leaves of greenhouse grown plants that were 2

weeks old. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a NotI restrictions site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electro- Max DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana- Champaign. email: l-vodkin@uiuc.edu"
BASE COUNT 123 a 67 c 92 g 106 t
ORIGIN

Query Match 72.0%; Score 18; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 cgcaagtcacagaattat 23
||||||||||||||||||
Db 334 CGCAAGTCACAGAATTAT 351

RESULT 3
BE658717/c
LOCUS BE658717 751 bp mRNA EST 24-MAY-2001
DEFINITION GM700007A10F9 Gm-r1070 Glycine max cDNA clone Gm-r1070-2561 3', mRNA sequence.
ACCESSION BE658717
VERSION BE658717.1 GI:9984609
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 751)
AUTHORS Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
TITLE A Functional Genomics Program for Soybean (NSF 9872565)
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: AW156684 corresponding to Gm-c1015-2632 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact:clones@genomesystems.com or info@genomesystems.com web site:www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES Location/Qualifiers
source 1. .751
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="Gm-r1070-2561"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, reracked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were reracked to form library Gm-r1070. The cDNA clones of the reracked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html . Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."
BASE COUNT 193 a 171 c 131 g 226 t 30 others
ORIGIN

Query Match 72.0%; Score 18; DB 10; Length 751;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 cgcaagtcacagaattat 23
||||||||||||||||||
Db 112 CGCAAGTCACAGAATTAT 95

RESULT 4
AQ227679
LOCUS AQ227679 424 bp DNA GSS 26-SEP-1998
DEFINITION HS_2012_B2_F10_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2012 Col=20 Row=L, DNA sequence.
ACCESSION AQ227679
VERSION AQ227679.1 GI:3652908
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 424)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2012 row: L column: 20
Class: BAC ends
High quality sequence stop: 424.
FEATURES Location/Qualifiers
source 1. .424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2012 Col=20 Row=L"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 107 a 86 c 87 g 143 t 1 others
ORIGIN

```
  Query Match             68.0%;  Score 17;  DB 13;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 9.6;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaattattt 25
          |||||||||||||||||
Db    250 AAGTCACAGAATTATTT 266


RESULT   5
AI529608/c
LOCUS       AI529608      944 bp    mRNA            EST       18-MAR-1999
DEFINITION  ui80b10.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1888699 5' similar to gb:L11650 Mouse cytochrome P-450 IIE1
            mRNA, complete cds (MOUSE);, mRNA sequence.
ACCESSION   AI529608
VERSION     AI529608.1  GI:4443743
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 944)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:973023
            Seq primer: custom primer used
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .944
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1888699"
                     /clone_lib="Sugano mouse liver mlia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      253 a    238 c    226 g    227 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 944;
  Best Local Similarity   100.0%;  Pred. No. 9.6;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 aagtcacagaattattt 25
          |||||||||||||||||
Db    906 AAGTCACAGAATTATTT 890


RESULT   6
BB141316/c
LOCUS       BB141316      263 bp    mRNA            EST       28-JUN-2000
DEFINITION  BB141316 RIKEN full-length enriched, adult female vagina Mus
            musculus cDNA clone 9930007O04 3', mRNA sequence.
ACCESSION   BB141316
VERSION     BB141316.1  GI:8796253
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 263)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
            ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
             Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
            ,Y. and Hayashizaki,Y.
             Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
             Carninci,P. and Hayashizaki,Y.
             High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
             Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .263
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="9930007O04"
                     /clone_lib="RIKEN full-length enriched, adult female
                     vagina"
                     /sex="female"
                     /tissue_type="vagina"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
```

contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 69 a 56 c 43 g 95 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 263;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacagaattat 23
||||||||||||||||
Db 62 CAAGTCACAGAATTAT 47

RESULT 7
BE181705/c
LOCUS BE181705 270 bp mRNA EST 22-JUN-2000
DEFINITION QV1-HT0638-080500-197-d06 HT0638 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE181705
VERSION BE181705.1 GI:8660881
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 270)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV1-HT0638-080500-197-d06&t3=2000-05-08&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 111
High quality sequence stop: 223.
FEATURES Location/Qualifiers
source 1. .270
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0638"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: pucl8; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 89 a 57 c 53 g 71 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 270;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 aagtcacagaattatt 24
|||||||||||||||||
Db 227 AAGTCACAGAATTATT 212

RESULT 8
BB450863
LOCUS BB450863 299 bp mRNA EST 21-JUL-2000
DEFINITION BB450863 RIKEN full-length enriched, 12 days embryo spinal ganglion Mus musculus cDNA clone D130006117 3', mRNA sequence.
ACCESSION BB450863
VERSION BB450863.1 GI:9311898
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 299)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .299
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D130006117"
/clone_lib="RIKEN full-length enriched, 12 days embryo spinal ganglion"

/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 75 a 72 c 45 g 107 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 299;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 cgcaagtcacagaatt 21
||||||||||||||||
Db 3 CGCAAGTCACAGAATT 18

RESULT 9
AA648449
LOCUS AA648449 353 bp mRNA EST 13-NOV-1997
DEFINITION ns22f10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1184395 3' similar to contains element LTR5 repetitive element ;, mRNA sequence.
ACCESSION AA648449
VERSION AA648449.1 GI:2574878
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 353)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
unknown library type
Insert Length: 2027 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 334.
FEATURES Location/Qualifiers
source 1. .353
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1184395"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 131 a 49 c 39 g 134 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 353;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 caagtcacagaattat 23
||||||||||||||||
Db 249 CAAGTCACAGAATTAT 264

RESULT 10
BF726110
LOCUS BF726110 405 bp mRNA EST 05-JAN-2001
DEFINITION by01c01.x1 Human Lens cDNA (Un-normalized, unamplified): BY Homo sapiens cDNA clone by01c01 3', mRNA sequence.
ACCESSION BF726110
VERSION BF726110.1 GI:12042021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 405)
AUTHORS Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics
JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 01 row: c column: 01
Seq primer: -21M13 forward primer (ABI).
FEATURES Location/Qualifiers
source 1. .405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="by01c01"
/clone_lib="Human Lens cDNA (Un-normalized, unamplified): BY"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMDH10B"
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies , essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's Instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adapter [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT 140 a 58 c 47 g 160 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 405;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db    249 CAAGTCACAGAATTAT 264


RESULT  11
BF232119
LOCUS       BF232119      429 bp    mRNA            EST       11-DEC-2000
DEFINITION  dd99d06.x1 Wellcome CRC pRN3 oocyte Xenopus laevis cDNA clone
            IMAGE:3436475 3', mRNA sequence.
ACCESSION   BF232119
VERSION     BF232119.1  GI:11174440
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
            Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
            University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1. .429
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:3436475"
                     /clone_lib="Wellcome CRC pRN3 oocyte"
                     /tissue_type="oocyte"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to  Nieuwkoop and Faber. Library was constructed
                     by N. Garrett, P. LeMaire, A.M. Zorn,  and J.B. Gurdon
                     (Wellcome/CRC Institute). "
BASE COUNT      141 a     73 c     74 g    141 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 429;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 agtcacagaattattt 25
          ||||||||||||||||
Db     60 AGTCACAGAATTATTT 75


RESULT  12
AI149091
LOCUS       AI149091      443 bp    mRNA            EST       28-OCT-1998
DEFINITION  qc85f02.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:1721019 3', mRNA sequence.
ACCESSION   AI149091
VERSION     AI149091.1  GI:3677560
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 527   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1721019"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT      126 a    122 c     71 g    124 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 443;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 agtcacagaattattt 25
          ||||||||||||||||
Db    100 AGTCACAGAATTATTT 115


RESULT  13
AQ086771
LOCUS       AQ086771      454 bp    DNA             GSS       26-AUG-1998
DEFINITION  HS_2175_B2_A06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2175 Col=12 Row=B, DNA sequence.
ACCESSION   AQ086771
VERSION     AQ086771.1  GI:3455988
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
```

```
            Plate: 2175  row: B  column: 12
            Class: BAC ends
            High quality sequence stop: 454.
FEATURES              Location/Qualifiers
     source           1. .454
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=2175 Col=12 Row=B"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /sex="male"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
BASE COUNT      188 a     70 c     76 g    117 t      3 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 454;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db    156 CAAGTCACAGAATTAT 171


RESULT  14
AI361386
LOCUS       AI361386      457 bp    mRNA            EST       15-FEB-1999
DEFINITION  qy47f01.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2015161 3',
            mRNA sequence.
ACCESSION   AI361386
VERSION     AI361386.1  GI:4113007
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1228   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 434.
FEATURES              Location/Qualifiers
     source           1. .457
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2015161"
                      /clone_lib="NCI_CGAP_Brn23"
                      /tissue_type="glioblastoma (pooled)"
                      /lab_host="DH10B"
                      /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                      modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5'
                      TGTTACCAATCTGAAGTGGGAGCGGCCGCATATCTTTTTTTTTTTTTTTTTTTTTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI
                      adaptors (Pharmacia), digested with Not I and cloned into
                      the Not I and Eco RI sites of the modified pT7T3 vector.
                      Library is normalized, and was constructed by Bento
                      Soares and M.Fatima Bonaldo."
BASE COUNT      153 a     73 c     52 g    179 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 457;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db    153 CAAGTCACAGAATTAT 168


RESULT  15
BG360303
LOCUS       BG360303      467 bp    mRNA            EST       06-MAR-2001
DEFINITION  dab93h11.y1 NICHD XGC Emb2 Xenopus laevis cDNA clone IMAGE:4405125
            5', mRNA sequence.
ACCESSION   BG360303
VERSION     BG360303.1  GI:13240994
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Life Technologies. DNA Sequencing by:
            Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 465.
FEATURES              Location/Qualifiers
     source           1. .467
                      /organism="Xenopus laevis"
                      /db_xref="taxon:8355"
                      /clone="IMAGE:4405125"
                      /clone_lib="NICHD XGC Emb2"
                      /tissue_type="embryo, stage 17/19"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                      Cloned unidirectionally. Primer: Oligo dT. Average insert
                      size 2.1 kb. Constructed by Life Technologies."
BASE COUNT      146 a     86 c     84 g    151 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 467;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 agtcacagaattattt 25
          ||||||||||||||||
Db     89 AGTCACAGAATTATTT 104


RESULT  16
AI635259
```

```
LOCUS       AI635259      471 bp    mRNA            EST       16-DEC-1999
DEFINITION  tz79h07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2294845 3',
            mRNA sequence.
ACCESSION   AI635259
VERSION     AI635259.1  GI:4686589
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Life Technologies catalog #: 11548-013
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1533    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2294845"
                     /clone_lib="NCI_CGAP_Pan1"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.72 kb. Life Technologies catalog #:
                     11548-013"
BASE COUNT      171 a     69 c     53 g    178 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 471;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db    235 CAAGTCACAGAATTAT 250


RESULT  17
AI634722
LOCUS       AI634722      503 bp    mRNA            EST       16-DEC-1999
DEFINITION  tz09e04.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2288094 3'
            similar to contains element XTR repetitive element ;, mRNA
            sequence.
ACCESSION   AI634722
VERSION     AI634722.1  GI:4686052
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 3719    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 413.
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2288094"
                     /clone_lib="NCI_CGAP_Ut1"
                     /tissue_type="well-differentiated endometrial
                     adenocarcinoma, 7 pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb. Life Technologies catalog #:
                     11538-014"
BASE COUNT      184 a     77 c     56 g    186 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 503;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaattat 23
          ||||||||||||||||
Db    236 CAAGTCACAGAATTAT 251


RESULT  18
BG347513
LOCUS       BG347513      503 bp    mRNA            EST       28-FEB-2001
DEFINITION  dab87h12.x1 NICHD XGC Emb2 Xenopus laevis cDNA clone IMAGE:4404719
            3', mRNA sequence.
ACCESSION   BG347513
VERSION     BG347513.1  GI:13167937
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Life Technologies. DNA Sequencing by:
            Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 432.
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4404719"
                     /clone_lib="NICHD XGC Emb2"
```

```
                     /tissue_type="embryo, stage 17/19"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.1 kb. Constructed by Life Technologies."
BASE COUNT      167 a     87 c     89 g    160 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 503;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           |||||||||||||||||
Db      51 AGTCACAGAATTATTT 66


RESULT  19
AW418528
LOCUS       AW418528      512 bp    mRNA            EST       09-FEB-2000
DEFINITION  xr32c07.x1 NCI_CGAP_Ut4 Homo sapiens cDNA clone IMAGE:2761836 3'
            similar to contains element MER32 repetitive element ;, mRNA
            sequence.
ACCESSION   AW418528
VERSION     AW418528.1  GI:6946410
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 404.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2761836"
                     /clone_lib="NCI_CGAP_Ut4"
                     /tissue_type="serous papillary carcinoma, high grade, 2
                     pooled tumors"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.48 kb. Life Technologies catalog #:
                     11542-016"
BASE COUNT      186 a     79 c     56 g    191 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 512;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db     240 CAAGTCACAGAATTAT 255


RESULT  20
AQ472309
LOCUS       AQ472309      527 bp    DNA             GSS       23-APR-1999
DEFINITION  CITBI-E1-2587F18.TF CITBI-E1 Homo sapiens genomic clone 2587F18,
            DNA sequence.
ACCESSION   AQ472309
VERSION     AQ472309.1  GI:4655963
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2587F18.TR
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2587F18"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT      150 a     88 c     83 g    205 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 527;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db      81 AGTCACAGAATTATTT 96


RESULT  21
BF322232/c
LOCUS       BF322232      649 bp    mRNA            EST       21-NOV-2000
DEFINITION  maa34h05.y1 NCI_CGAP_Li10 Mus musculus cDNA clone IMAGE:3812985 5'
            similar to SW:CPE1_MOUSE Q05421 CYTOCHROME P450 2E1 ;, mRNA
            sequence.
ACCESSION   BF322232
VERSION     BF322232.1  GI:11271563
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 649)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
```

```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov

            MGI:1455097
            Seq primer: -40RP from Gibco
            High quality sequence stop: 434.
FEATURES              Location/Qualifiers
     source           1. .649
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3812985"
                      /clone_lib="NCI_CGAP_Li10"
                      /sex="female"
                      /dev_stage="10 weeks"
                      /lab_host="DH10B (T1 phage-resistant)"
                      /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                      Average insert size 1.6 kb.   Library constructed by Life
                      Technologies."
BASE COUNT      201 a    150 c    150 g    147 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 649;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db     517 AGTCACAGAATTATTT 502


RESULT  22
AQ541562/c
LOCUS       AQ541562      652 bp    DNA             GSS       19-MAY-1999
DEFINITION  RPCI-11-361M22.TV RPCI-11 Homo sapiens genomic clone RPCI-11-361M22
            , DNA sequence.
ACCESSION   AQ541562
VERSION     AQ541562.1  GI:4872018
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-361M22.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .652
                      /organism="Homo sapiens"
                      /db_xref="GDB:7638549"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-361M22"
                      /clone_lib="RPCI-11"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
BASE COUNT      201 a    162 c     82 g    207 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 652;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db      96 CAAGTCACAGAATTAT 81


RESULT  23
BF726111/c
LOCUS       BF726111      657 bp    mRNA            EST       05-JAN-2001
DEFINITION  by01c01.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
            sapiens cDNA clone by01c01 5', mRNA sequence.
ACCESSION   BF726111
VERSION     BF726111.1  GI:12042022
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
  TITLE     NEIBANK: EST analysis and bioinformatics for ocular genomics
  JOURNAL   Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT     Contact: Wistow G
            Section on Molecular Structure and Function
            National Eye Institute
            6/331, NIH, Bethesda, MD 20892-2740, USA
            Tel: 301 402 3452
            Fax: 301 496 0078
            Email: graeme@helix.nih.gov
            Plate: 01  row: c  column: 01
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES              Location/Qualifiers
     source           1. .657
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="by01c01"
                      /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
                      BY"
                      /tissue_type="Lens"
                      /dev_stage="Adult"
                      /lab_host="EMDH10B"
                      /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
                      from different adults (both approximately 40 years old)
                      together yielded 20ug of total RNA and 150ng mRNA for cDNA
                      library synthesis. A directionally cloned cDNA library in
                      the pCMVSPORT6 vector was constructed at Life Technologies
                      , essentially following the protocols of the SuperScript
                      Plasmid System full details of which are contained in the
                      manufacturer's Instruction manual
                      (http://www.lifetech.com/). First strand synthesis was
                      carried out using a Not I primer-adapter
                      [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)15-3']. Not I/blunt
                      end inserts were cloned into the Not I/EcoR V sites in the
                      vector. EST analysis was performed on the unamplified
                      library at the NIH Intramural Sequencing Center (NISC)."
```

```
BASE COUNT      232 a     86 c    114 g    225 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 657;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaattat 23
           ||||||||||||||||
Db     574 CAAGTCACAGAATTAT 559


RESULT  24
AZ382380
LOCUS       AZ382380      702 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0139H04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0139H04 R, DNA sequence.
ACCESSION   AZ382380
VERSION     AZ382380.1  GI:10496080
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 702)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0139  row: H  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 702.
FEATURES             Location/Qualifiers
     source          1. .702
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0139H04"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      272 a    120 c    104 g    206 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 702;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattatt 24
           ||||||||||||||||
Db     637 AAGTCACAGAATTATT 652


RESULT  25
BG496183/c
LOCUS       BG496183      753 bp    mRNA            EST       27-MAR-2001
DEFINITION  602538125F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:4658982 5',
            mRNA sequence.
ACCESSION   BG496183
VERSION     BG496183.1  GI:13457699
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1454  row: b  column: 07
            High quality sequence start: 20
            High quality sequence stop: 694.
FEATURES             Location/Qualifiers
     source          1. .753
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4658982"
                     /clone_lib="NIH_MGC_59"
                     /tissue_type="mucoepidermoid carcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      202 a    143 c    174 g    233 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 aagtcacagaattatt 24
           ||||||||||||||||
Db      46 AAGTCACAGAATTATT 31
```

```
RESULT  26
AQ900023/c
LOCUS       AQ900023      971 bp    DNA             GSS       10-NOV-1999
DEFINITION  HS_3180_A2_G11_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3180 Col=22 Row=M, DNA sequence.
ACCESSION   AQ900023
VERSION     AQ900023.1  GI:6356213
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 971)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3180  row: M  column: 22
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 971.
FEATURES             Location/Qualifiers
     source          1. .971
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3180 Col=22 Row=M"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      249 a    130 c    128 g    464 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 971;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 agtcacagaattattt 25
           ||||||||||||||||
Db     501 AGTCACAGAATTATTT 486


RESULT  27
BG871816/c
LOCUS       BG871816     1318 bp    mRNA            EST       29-MAY-2001
DEFINITION  602792574F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4923619 5',
            mRNA sequence.
ACCESSION   BG871816
VERSION     BG871816.1  GI:14222356
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1318)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10845  row: d  column: 20
            High quality sequence start: 3
            High quality sequence stop: 211.
FEATURES             Location/Qualifiers
     source          1. .1318
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4923619"
                     /clone_lib="NCI_CGAP_SG2"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo
                     dT. Average insert size 1.3 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      379 a    339 c    297 g    303 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 1318;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 attgtcgcaagtcaca 16
             ||||||||||||||||
Db    1084 ATTGTCGCAAGTCACA 1069


RESULT  28
BG977588
LOCUS       BG977588     172 bp    mRNA            EST       12-JUN-2001
DEFINITION  PM4-CI0084-040101-006-d11 CI0084 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG977588
VERSION     BG977588.1  GI:14380323
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=PM4&t2=PM4-CI0084-
            040101-006-d11&t3=2001-01-04&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 165.
```

```
FEATURES             Location/Qualifiers
     source          1. .172
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CI0084"
                     /dev_stage="Adult"
                     /note="Organ: colon_ins; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       37 a     51 c     55 g     28 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 172;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gtcgcaagtcacaga 18
           ||||||||||||||||
Db       8 GTCGCAAGTCACAGA 22


RESULT  29
AW871559/c
LOCUS       AW871559      214 bp    mRNA            EST       10-MAY-2001
DEFINITION  ra79d04.y1 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
            cDNA 5' similar to WP:C26D10.4 CE03027 ;, mRNA sequence.
ACCESSION   AW871559
VERSION     AW871559.1  GI:8005612
KEYWORDS    EST.
SOURCE      southern root-knot nematode.
  ORGANISM  Meloidogyne incognita
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 214)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.
            , Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.
            , Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R.
            , Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
            ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.
            , Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
            Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            The library was constructed by Uma Rao and David Bird
            (david_bird@ncsu.edu) at North Carolina State University. DNA
            Sequencing by: Washington University Genome Sequencing Center St.
            Louis.
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            Seq primer: -40RP from Gibco
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .214
                     /organism="Meloidogyne incognita"
                     /db_xref="taxon:6306"
                     /clone_lib="Bird-Rao Meloidogyne incognita J2"
                     /dev_stage="enriched for 2nd stage juveniles"
                     /lab_host="XLORL"
                     /note="Vector: ZAP express - pBKCMV (Stratagene); Site_1:
                     EcoRI; Site_2: XhoI; Oligo (dT) primed library.  cDNA was
                     contructed and cloned  unidirectionally into the vector
                     within the 5' EcoRI and  3' XhoI sites.  This library was
                     constructed by Dr. Uma Rao  and Dr. David Bird at North
                     Carolina state University."
BASE COUNT       59 a     52 c     19 g     84 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 214;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db     184 CAAGTCACAGAATTA 170


RESULT  30
BE763893
LOCUS       BE763893      233 bp    mRNA            EST       19-SEP-2000
DEFINITION  RC4-NT0054-120600-015-g05 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE763893
VERSION     BE763893.1  GI:10193817
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 233)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC4-NT0054-120
            600-015-g05&t3=2000-06-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 233.
FEATURES             Location/Qualifiers
     source          1. .233
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0054"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       88 a     27 c     52 g     66 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 233;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db      94 CAAGTCACAGAATTA 108


RESULT  31
BE763892
LOCUS       BE763892      253 bp    mRNA            EST       19-SEP-2000
DEFINITION  RC4-NT0054-120600-015-g03 NT0054 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE763892
VERSION     BE763892.1  GI:10193816
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 253)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=RC4-NT0054-120
            600-015-g03&t3=2000-06-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 12
            High quality sequence stop: 253.
FEATURES             Location/Qualifiers
     source          1. .253
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NT0054"
                     /dev_stage="Adult"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       95 a     31 c     56 g     71 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 253;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db     114 CAAGTCACAGAATTA 128


RESULT  32
AA317110
LOCUS       AA317110      256 bp    mRNA            EST       19-APR-1997
DEFINITION  EST19001 Lung Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA317110
VERSION     AA317110.1  GI:1969448
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 256)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC99418
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .256
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):117571"
                     /db_xref="taxon:9606"
                     /clone_lib="Lung"
                     /dev_stage="adult"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI"
BASE COUNT       73 a     57 c     48 g     75 t      3 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 256;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db     189 CAAGTCACAGAATTA 203


RESULT  33
Z45844/c
LOCUS       Z45844        261 bp    mRNA            EST       14-NOV-1994
DEFINITION  HSCZVE041 normalized infant brain cDNA Homo sapiens cDNA clone
            c-zve04, mRNA sequence.
ACCESSION   Z45844
VERSION     Z45844.1  GI:575078
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
            ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
            Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: ylc-zve04
            Seq primer: (-21)M13_universal.
FEATURES             Location/Qualifiers
     source          1. .261
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="c-zve04"
                     /clone_lib="normalized infant brain cDNA"
                     /sex="Female"
                     /tissue_type="total brain"
                     /dev_stage="3 months old"
                     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                     Site_2: NotI; sex=Female; dev_stage=3 months old;
                     isolate=muscular atrophy patient;  tissue_type=total brain
                     ;  total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the  HindIII -> NotI sites of the
                     lafmid BA vector. Clone library from B.Soares,  Psychiatry
                     Dept. Columbia University, USA.  Normalization_method:
                     Bento Soares, P.N.A.S in press"
BASE COUNT       77 a     56 c     31 g     97 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 261;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           ||||||||||||||||
Db     138 CAAGTCACAGAATTA 124


RESULT  34
BF745516
LOCUS       BF745516      263 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-021100-469-h05 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF745516
VERSION     BF745516.1  GI:12072192
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 263)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            021100-469-h05&t3=2000-11-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 129
            High quality sequence stop: 263.
FEATURES             Location/Qualifiers
     source          1. .263
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0826"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: pucl8; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       77 a     55 c     60 g     71 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 263;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaatt 21
           |||||||||||||||
Db     217 GCAAGTCACAGAATT 231


RESULT  35
AA115083/c
LOCUS       AA115083      266 bp    mRNA            EST       14-MAY-1997
DEFINITION  zl06c10.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:491538 5', mRNA sequence.
ACCESSION   AA115083
VERSION     AA115083.1  GI:1670318
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 266)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 555   Std Error: 0.00
            Seq primer: -28M13 rev2 from Amersham.
FEATURES             Location/Qualifiers
```

```
     source          1. .266
                     /organism="Homo sapiens"
                     /db_xref="GDB:3806125"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:491538"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not 1;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT       61 a     67 c     64 g     69 t      5 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 266;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 gtcacagaattattt 25
           |||||||||||||||
Db      91 GTCACAGAATTATTT 77


RESULT  36
BF800965/c
LOCUS       BF800965      282 bp    mRNA            EST       12-JAN-2001
DEFINITION  QV4-CI0100-231000-457-d12 CI0100 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF800965
VERSION     BF800965.1  GI:12129954
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 282)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV4&t2=QV4-CI0100-
            231000-457-d12&t3=2000-10-23&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 59
            High quality sequence stop: 282.
FEATURES             Location/Qualifiers
     source          1. .282
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CI0100"
                     /dev_stage="Adult"
                     /note="Organ: colon_ins; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       68 a     75 c     67 g     72 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 282;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 attgtcgcaagtcac 15
          |||||||||||||||
Db     51 ATTGTCGCAAGTCAC 37


RESULT  37
AA344205
LOCUS       AA344205      286 bp    mRNA            EST       21-APR-1997
DEFINITION  EST50086 Gall bladder I Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA344205
VERSION     AA344205.1  GI:1996443
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 286)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC99418
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .286
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):145738"
                     /db_xref="taxon:9606"
                     /clone_lib="Gall bladder I"
                     /sex="female"
```

```
                     /dev_stage="adult, 25 yrs"
                     /note="Organ: gall bladder; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       83 a     62 c     48 g     91 t      2 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 286;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatta 22
          ||||||||||||||||
Db    197 CAAGTCACAGAATTA 211


RESULT  38
B53227/c
LOCUS       B53227         292 bp    DNA             GSS       20-JUN-1998
DEFINITION  CIT-HSP-2009H2.TR CIT-HSP Homo sapiens genomic clone 2009H2, DNA
            sequence.
ACCESSION   B53227
VERSION     B53227.1  GI:2607561
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
            and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Homo sapiens"
                     /db_xref="GDB:7040812"
                     /db_xref="taxon:9606"
                     /clone="2009H2"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT       58 a     64 c     90 g     80 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatta 22
          ||||||||||||||||
Db    248 CAAGTCACAGAATTA 234


RESULT  39
BF743951/c
LOCUS       BF743951       317 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-c02 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743951
VERSION     BF743951.1  GI:12070627
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            181000-428-c02&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 17
            High quality sequence stop: 317.
FEATURES             Location/Qualifiers
     source          1. .317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BT0826"
                     /dev_stage="Adult"
                     /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       96 a     71 c     49 g    101 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 317;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaatt 21
          |||||||||||||||
Db     45 GCAAGTCACAGAATT 31


RESULT  40
BH121038/c
LOCUS       BH121038       324 bp    DNA             GSS       19-JUL-2001
DEFINITION  RPCI-24-230E13.TV RPCI-24 Mus musculus genomic clone RPCI-24-230E13
            , DNA sequence.
ACCESSION   BH121038
VERSION     BH121038.1  GI:14964550
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 324)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-230E13.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 230  row: E  column: 13
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .324
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-230E13"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT       93 a     57 c     88 g     86 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 324;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gcaagtcacagaatt 21
           |||||||||||||||
Db      26 GCAAGTCACAGAATT 12


RESULT  41
T30901/c
LOCUS       T30901          327 bp    mRNA            EST       06-SEP-1995
DEFINITION  EST24536 Human Brain Homo sapiens cDNA 5' end similar to None, mRNA
            sequence.
ACCESSION   T30901
VERSION     T30901.1  GI:612999
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 327)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.-W., Clayton,R.A.,
            Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D., FitzGerald
            ,L.M., FitzHugh,W.M., Fritchman,J.L., Geoghagen,N.S.M., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., HinkleJr,P.S., Kelley,J.M.,
            Klimek,K.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pellegrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.-J.,
            Dimke,D., Feng,P., Ferrie,A., Fischer,C., Hastings,G.A., He,W.-W.,
            Hu,J.-S., Greene,J.M., Gruber,J., Hudson,P., Kim,A., Kozak,D.L.,
            Kunsch,C., Ji,H., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei
            ,Y.-F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial Assessment of Human Gene Diversity and Expression Patterns
            Based Upon 83 Million Basepairs of cDNA Sequence
  JOURNAL   Nature 377, 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC10632
            Contact: Venter, JC
            The Institute for Genomic Research
            932 Clopper Rd, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: tdbinfo@tdb.tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please contact the TIGR Database
            (tdbinfo@tdb.tigr.org)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .327
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):100882"
                     /db_xref="taxon:9606"
                     /clone_lib="Human Brain"
                     /note="Organ: brain"
BASE COUNT       89 a     75 c     47 g    115 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 caagtcacagaatta 22
           |||||||||||||||
Db     295 CAAGTCACAGAATTA 281


RESULT  42
BF743964/c
LOCUS       BF743964        335 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-h04 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743964
VERSION     BF743964.1  GI:12070640
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 335)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
```

```
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            181000-428-h04&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 40
            High quality sequence stop: 335.
FEATURES              Location/Qualifiers
     source           1. .335
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="BT0826"
                      /dev_stage="Adult"
                      /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT      101 a     76 c     55 g    103 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 335;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaatt 21
          |||||||||||||||
Db     61 GCAAGTCACAGAATT 47


RESULT  43
BF743954/c
LOCUS       BF743954       346 bp    mRNA            EST       10-JAN-2001
DEFINITION  CM2-BT0826-181000-428-d07 BT0826 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF743954
VERSION     BF743954.1  GI:12070630
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 346)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=CM2&t2=CM2-BT0826-
            181000-428-d07&t3=2000-10-18&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 43
            High quality sequence stop: 346.
FEATURES              Location/Qualifiers
     source           1. .346
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="BT0826"
                      /dev_stage="Adult"
                      /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No. 196
                      ,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
BASE COUNT      103 a     78 c     54 g    111 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 346;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 gcaagtcacagaatt 21
          |||||||||||||||
Db     73 GCAAGTCACAGAATT 59


RESULT  44
T07051
LOCUS       T07051         370 bp    mRNA            EST       30-JUN-1993
DEFINITION  EST04940 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
            clone HFBEC79, mRNA sequence.
ACCESSION   T07051
VERSION     T07051.1  GI:318200
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     3,400 expressed sequence tags identify diversity of transcripts
            from human brain
  JOURNAL   Nature Genet. 4, 256-267 (1993)
  MEDLINE   93364420
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13-21.
FEATURES              Location/Qualifiers
     source           1. .370
                      /organism="Homo sapiens"
                      /db_xref="ATCC (inhost):83698"
                      /db_xref="taxon:9606"
                      /clone="HFBEC79"
                      /clone_lib="Fetal brain, Stratagene (cat#936206)"
                      /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
                      oligo-dT + random primed cDNA synthesis; lambdaZAP-II
                      vector, 1.0kb average inser size."
BASE COUNT      137 a     58 c     68 g    106 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 370;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 caagtcacagaatta 22
          |||||||||||||||
Db    113 CAAGTCACAGAATTA 127


RESULT  45
H53286/c
LOCUS       H53286         381 bp    mRNA            EST       20-SEP-1995
DEFINITION  yq85f09.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
```

```
            IMAGE:202601 5', mRNA sequence.
ACCESSION   H53286
VERSION     H53286.1  GI:993433
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 381)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1459
            High quality sequence stops: 288
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1459   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 288.
FEATURES             Location/Qualifiers
     source          1. .381
                     /organism="Homo sapiens"
                     /db_xref="GDB:3771732"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:202601"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      105 a     99 c     79 g     88 t     10 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 381;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 gtcacagaattattt 25
           |||||||||||||||
Db     281 GTCACAGAATTATTT 267



Search completed: December 26, 2001, 14:09:56
Job time: 7043 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:50:42 ; Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: US-09-396-196F-3
Perfect score: 25
Sequence: 1 gttgctggatctgctgtttgaagcg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 270703

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		25	100.0	128	1	ECOBIOB	M27731 Escherichia
2		25	100.0	1041	6	AR029499	AR029499 Sequence
3		25	100.0	1041	6	AR034916	AR034916 Sequence
4		25	100.0	1084	6	A11530	A11530 BioB gene o
5		25	100.0	5526	1	AF250776	AF250776 Unculture
6		25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
7		25	100.0	5872	6	A38246	A38246 Sequence 1
8		25	100.0	5872	6	A38251	A38251 Sequence 6
9		25	100.0	5872	6	A93674	A93674 Sequence 1
10		25	100.0	5872	6	A93679	A93679 Sequence 6
11		25	100.0	5872	6	AR101809	AR101809 Sequence
12		25	100.0	5872	6	AR101810	AR101810 Sequence
13		25	100.0	11022	1	AE000180	AE000180 Escherich
14		25	100.0	13501	1	AE005258	AE005258 Escherich
15		25	100.0	297816	1	AP002553	AP002553 Escherich
16		24	96.0	1121	6	E00893	E00893 Genomic DNA
17		20	80.0	12891	1	AE004192	AE004192 Vibrio ch
18	c	17	68.0	383	3	S64534	S64534 Drosophila
19	c	17	68.0	461	3	S64532	S64532 Drosophila
20	c	17	68.0	516	3	DMDECG1B	X76231 D.melanogas
21	c	17	68.0	539	3	S64530	S64530 Drosophila
22	c	17	68.0	539	3	S64531	S64531 Drosophila
23	c	17	68.0	617	3	S64529	S64529 Drosophila
24	c	17	68.0	774	3	DSDEC1TR6	X96931 D.sechellia
25	c	17	68.0	852	3	DSDEC1TR2	X96927 D.simulans
26	c	17	68.0	852	3	DSDEC1TR3	X96928 D.simulans
27	c	17	68.0	930	3	DMDEC1TR1	X96926 D.melanogas
28	c	17	68.0	3741	3	DRODEC1A	M35887 Drosophila
29	c	17	68.0	3755	3	DRODEC1B	M35888 Drosophila
30	c	17	68.0	5697	3	DRODEC1C	M35889 Drosophila
31	c	17	68.0	33204	2	AC017995	AC017995 Drosophil
32	c	17	68.0	161540	2	AC016952	AC016952 Homo sapi
33		17	68.0	306267	3	AE003442	AE003442 Drosophil
34	c	16	64.0	774	3	DMDEC1TR5	X96930 D.mauritian
35	c	16	64.0	1044	5	AF250347	AF250347 Xenopus l
36	c	16	64.0	1614	5	AF281080	AF281080 Xenopus l
37		16	64.0	5000	6	AR120145	AR120145 Sequence
38		16	64.0	5000	8	AF143940	AF143940 Arabidops
39		16	64.0	6887	8	YLI012084	AJ012084 Yarrowia
40	c	16	64.0	11328	1	AE004322	AE004322 Vibrio ch
41		16	64.0	13856	1	AE004026	AE004026 Xylella f
42		16	64.0	15247	1	AE004580	AE004580 Pseudomon
43		16	64.0	31557	9	HS314G4	Z69667 Human DNA s
44		16	64.0	41200	8	SC8520X	Z49705 S.cerevisia
45	c	16	64.0	46340	6	AX008043	AX008043 Sequence
46		16	64.0	62927	8	AB018119	AB018119 Arabidops
47	c	16	64.0	65679	2	AC013956	AC013956 Drosophil
48	c	16	64.0	74127	2	AC020125	AC020125 Drosophil
49		16	64.0	78529	2	AC012807	AC012807 Drosophil
50	c	16	64.0	79048	2	AC008298	AC008298 Drosophil
51	c	16	64.0	102540	8	AC079281	AC079281 Arabidops
52		16	64.0	109683	2	AC073020	AC073020 Homo sapi
53		16	64.0	110000	2	LMFLCHR31_15	Continuation (16 o
54		16	64.0	150594	8	AP001859	AP001859 Oryza sat
55	c	16	64.0	156425	8	AP002897	AP002897 Oryza sat
56		16	64.0	158148	2	AC015463	AC015463 Homo sapi
57		16	64.0	158560	2	AC073357	AC073357 Homo sapi
58		16	64.0	166919	2	AL591675	AL591675 Mus muscu
59		16	64.0	167862	9	AC011966	AC011966 Homo sapi
60	c	16	64.0	168177	2	AC023914	AC023914 Homo sapi
61	c	16	64.0	169893	2	AC007532	AC007532 Drosophil
62		16	64.0	171069	2	AP003845	AP003845 Oryza sat
63	c	16	64.0	177163	9	HS134E15	AL022067 Human DNA
64		16	64.0	177857	3	AC010211	AC010211 Drosophil
65		16	64.0	177989	2	AL591430	AL591430 Mus muscu
66		16	64.0	187272	2	AC069076	AC069076 Homo sapi
67		16	64.0	188055	2	AC023593	AC023593 Homo sapi
68		16	64.0	188156	2	AC079277	AC079277 Mus muscu
69	c	16	64.0	189252	2	AC018930	AC018930 Homo sapi
70		16	64.0	190029	2	AC068664	AC068664 Mus muscu

	71	16	64.0	191677	2	AC090534	AC090534 Mus muscu
c	72	16	64.0	200000	2	AC007104	AC007104 Homo sapi
	73	16	64.0	200900	9	AC013410	AC013410 Homo sapi
	74	16	64.0	214241	2	AC084163	AC084163 Mus muscu
	75	16	64.0	221420	2	AC011822	AC011822 Homo sapi
	76	16	64.0	229155	9	AC018755	AC018755 Homo sapi
	77	16	64.0	260418	9	AE006463	AE006463 Homo sapi
c	78	16	64.0	266344	9	AC005158	AC005158 Homo sapi
	79	16	64.0	277023	3	AE003519	AE003519 Drosophil
c	80	16	64.0	291545	3	AE003600	AE003600 Drosophil
	81	15	60.0	306	14	PPHE7D	D50548 Human papil
c	82	15	60.0	358	11	G00476	G00476 fruit fly S
c	83	15	60.0	503	3	DEDECG1B	X76233 D.erecta de
c	84	15	60.0	708	6	A86338	A86338 Sequence 99
c	85	15	60.0	708	6	AR155831	AR155831 Sequence
c	86	15	60.0	708	6	E66356	E66356 Genome DNA
	87	15	60.0	1041	1	EHBIOTOP2	U38648 Erwinia her
c	88	15	60.0	1172	8	ZMSNORNA2	Y11049 Z.mays smal
c	89	15	60.0	1616	10	AF239157	AF239157 Rattus no
c	90	15	60.0	1623	10	AF009246	AF009246 Mus muscu
c	91	15	60.0	1683	6	A72446	A72446 Sequence 33
c	92	15	60.0	1683	6	A80156	A80156 Sequence 33
	93	15	60.0	2018	3	AE002676	AE002676 Drosophil
c	94	15	60.0	2021	2	AC019735	AC019735 Drosophil
c	95	15	60.0	2546	9	AK026448	AK026448 Homo sapi
	96	15	60.0	3569	3	CEU20168	U20168 Caenorhabdi
	97	15	60.0	4116	1	PSEKDGDH	M69160 Pseudomonas
c	98	15	60.0	5607	3	DROPROSA	D10609 Drosophila
c	99	15	60.0	5869	9	AF181672	AF181672 Homo sapi
c	100	15	60.0	6017	3	DMPROSPER	Z11743 D. melanoga

ALIGNMENTS

```
RESULT   1
ECOBIOB
LOCUS       ECOBIOB       128 bp    DNA             BCT       03-JAN-1995
DEFINITION  Escherichia coli biotin (bioB) gene, early terminator region.
ACCESSION   M27731
VERSION     M27731.1  GI:341755
KEYWORDS    bioB gene; biotin.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 128)
  AUTHORS   Nath,S.K.
  TITLE     Attenuation of transcription of biotin genes in Escherichia coli
  JOURNAL   Can. J. Microbiol. 34 (12), 1288-1296 (1988)
  MEDLINE   89167942
FEATURES             Location/Qualifiers
     source          1. .128
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     gene            4. .128
                     /gene="bioB"
     -10_signal      4. .9
                     /gene="bioB"
     mRNA            20. .>128
                     /gene="bioB"
     terminator      103. .121
                     /gene="bioB"
                     /note="early terminator"
BASE COUNT       31 a      28 c      34 g      35 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 1;  Length 128;
  Best Local Similarity   100.0%;  Pred. No. 0.0019;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db      18 GTTGCTGGATCTGCTGTTTGAAGCG 42


RESULT   2
AR029499
LOCUS       AR029499      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT       262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0016;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db      57 GTTGCTGGATCTGCTGTTTGAAGCG 81


RESULT   3
AR034916
LOCUS       AR034916      1041 bp    DNA             PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT       262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.0016;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db      57 GTTGCTGGATCTGCTGTTTGAAGCG 81


RESULT   4
A11530
LOCUS       A11530        1084 bp    DNA             PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
```

```
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.0016;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db      80 GTTGCTGGATCTGCTGTTTGAAGCG 104


RESULT   5
AF250776
LOCUS       AF250776     5526 bp    DNA             BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
            modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
            biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
            biotin biosynthesis protein BioC (bioC) genes, complete cds; and
            dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION   AF250776
VERSION     AF250776.1  GI:12620124
KEYWORDS    .
SOURCE      uncultured bacterium pCosHE2.
  ORGANISM  uncultured bacterium pCosHE2
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct cloning from enrichment cultures, a reliable strategy for
            isolation of complete operons and genes from microbial consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE   20575196
   PUBMED   11133432
REFERENCE   2  (bases 1 to 5526)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .5526
                     /organism="uncultured bacterium pCosHE2"
                     /db_xref="taxon:143797"
                     /clone="pCosHE2"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     CDS             complement(52. .528)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical 17.1 kDa protein in modC-bioA
                     intergenic region"
                     /protein_id="AAG60577.1"
                     /db_xref="GI:12620125"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(587. .1876)
                     /gene="bioA"
     CDS             complement(587. .1876)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase BioA"
                     /protein_id="AAG60578.1"
                     /db_xref="GI:12620126"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            1963. .3003
                     /gene="bioB"
     CDS             1963. .3003
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG60579.1"
                     /db_xref="GI:12620127"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3000. .4154
                     /gene="bioF"
     CDS             3000. .4154
                     /gene="bioF"
                     /note="8-amino-7-oxononanoate synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA synthetase BioF"
                     /protein_id="AAG60580.1"
                     /db_xref="GI:12620128"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4141. .4896
                     /gene="bioC"
     CDS             4141. .4896
                     /gene="bioC"
                     /note="reaction step prior to synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60581.1"
                     /db_xref="GI:12620129"
                     /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
```

```
                     HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4883. .>5526
                     /gene="bioD"
     CDS             4883. .>5526
                     /gene="bioD"
                     /note="DTB synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase BioD"
                     /protein_id="AAG60582.1"
                     /db_xref="GI:12620130"
                     /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                     GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                     SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                     TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT     1274 a   1507 c   1567 g   1178 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 1;  Length 5526;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db   2019 GTTGCTGGATCTGCTGTTTGAAGCG 2043


RESULT   6
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAATHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
gene            2012. .3052
                /gene="bioB"
CDS             2012. .3052
                /gene="bioB"
                /codon_start=1
                /transl_table=11
                /product="biotin synthetase"
                /protein_id="AAA23515.1"
                /db_xref="GI:145425"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            3049. .4203
                /gene="bioF"
CDS             3049. .4203
                /gene="bioF"
                /codon_start=1
                /transl_table=11
                /product="7-keto-8-amino pelargonic acid synthetase"
                /protein_id="AAA23516.1"
                /db_xref="GI:145426"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
gene            4190. .4945
                /gene="bioC"
CDS             4190. .4945
                /gene="bioC"
                /codon_start=1
                /transl_table=11
                /protein_id="AAA23517.1"
                /db_xref="GI:145427"
                /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
                HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
gene            4938. .5597
                /gene="bioD"
CDS             4938. .5597
                /gene="bioD"
                /codon_start=1
                /transl_table=11
                /product="dethiobiotin synthetase"
                /protein_id="AAA23518.1"
                /db_xref="GI:145428"
                /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                T"
```

```
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match             100.0%;  Score 25;  DB 1;  Length 5793;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db   2068 GTTGCTGGATCTGCTGTTTGAAGCG 2092


RESULT   7
A38246
LOCUS       A38246       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /number=1
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAA02325.1"
                     /db_xref="GI:2294845"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANS1FYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=3
                     /product="PROTEIN"
                     /protein_id="CAA02326.1"
                     /db_xref="GI:2294846"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:
                     8-AMINO-7-OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /number=5
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAA02327.1"
                     /db_xref="GI:2294847"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /number=6
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAA02328.1"
                     /db_xref="GI:2294848"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197
```

```
RESULT   8
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"

BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity  100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT   9
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
```

```
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT  10
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT  11
AR101809
LOCUS       AR101809     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
```

```
          Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT  12
AR101810
LOCUS       AR101810     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
          Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.0014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db    173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT  13
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
          genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
          Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
          Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
          Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
          University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
          Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
          608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
          University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
          Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
          608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
          Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
          University of Wisconsin-Madison (Frederick R. Blattner, director).
          Supported by NIH grants HG00301 and HG01428 (from the Human Genome
          Project and NCHGR). The entire sequence was independently
          determined from E. coli K12 strain MG1655. Predicted open reading
          frames were determined using GeneMark software, kindly supplied by
          Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
          30332 [e-mail: mark@amber.gatech.edu].  Open reading frames that
          have been correlated with genetic loci are being annotated with CG
          Site Nos., unique ID nos. for the genes in the E. coli Genetic
          Stock Center (CGSC) database at Yale University, kindly supplied by
          Mary Berlyn. A public version of the database is accessible
          (http://cgsc.biology.yale.edu). Annotation of the genome is an
          ongoing task whose goal is to make the genome sequence more useful
          by correlating it with other data.  Comments to the authors are
          appreciated. Updated information will be available at the E. coli
          Genome Project's World Wide Web site
          (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
          its annotations are periodically updated; this is version M54. No
          sequence changes. Annotation updates: updated gene identifications
          and products; all new functional assignments courtesy of Monica
          Riley; added promoters, protein binding sites, and repeated
          sequences described in reference 1. The unique numeric identifiers
          beginning with a lowercase 'b' assigned to each gene (protein- or
          RNA-encoding) are now designated as gene synonyms instead of
          labels. This should allow them to be searched for in Entrez as gene
          names.
FEATURES      .      Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
     CDS             complement(147. .1430)
                     /gene="ybhC"
                     /function="orf; Not classified"
                     /note="f427; 98 pct identical to fragment YBHC_ECOLI
                     SW:P46130 (300 aa) but contains 127 additional C-terminal
                     residues"
                     /codon_start=1
                     /transl_table=11
                     /product="putative pectinesterase"
                     /protein_id="AAC73859.1"
                     /db_xref="GI:1786989"
                     /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                     LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                     AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                     DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                     LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
```

```
                NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
promoter        complement(1507. .1535)
                /note="factor Sigma70; predicted +1 start at 806574"
gene            complement(1582. .2058)
                /gene="ybhB"
                /note="b0773"
CDS             complement(1582. .2058)
                /gene="ybhB"
                /function="orf; Unknown"
                /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC73860.1"
                /db_xref="GI:1786990"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
protein_bind    complement(2108. .2124)
                /note="central position to predicted promoter:85.5"
                /bound_moiety="RhaS predicted site"
gene            complement(2117. .3406)
                /gene="bioA"
                /note="b0774"
CDS             complement(2117. .3406)
                /gene="bioA"
                /EC_number="2.6.1.62"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                /codon_start=1
                /transl_table=11
                /product="7,8-diaminopelargonic acid synthetase"
                /protein_id="AAC73861.1"
                /db_xref="GI:1786991"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
promoter        complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
protein_bind    3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
protein_bind    3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
protein_bind    complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
promoter        3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
promoter        complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
gene            3493. .4533
                /gene="bioB"
                /note="b0775"
CDS             3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
gene            4530. .5684
                /gene="bioF"
                /note="b0776"
CDS             4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
```

```
  Query Match             100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity   100.0%;  Pred. No. 0.0013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db   3549 GTTGCTGGATCTGCTGTTTGAAGCG 3573


RESULT  14
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
```

```
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EjbA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
```

```
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match             100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity   100.0%;  Pred. No. 0.0013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db   9635 GTTGCTGGATCTGCTGTTTGAAGCG 9659


RESULT  15
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
     gene            1245. .2114
                     /gene="ECs0754"
     CDS             1245. .2114
                     /gene="ECs0754"
                     /note="similar to SUCD_ECOLI gi|1786949 percent identity
                     100 in 289 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase alpha subunit"
                     /protein_id="BAB34177.1"
                     /db_xref="GI:13360213"
                     /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                     GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                     LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                     AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                     AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                     LADIGEALKTVLK"
     gene            2781. .3686
                     /gene="ECs0755"
     CDS             2781. .3686
                     /gene="ECs0755"
                     /note="probable transcriptional regulator,similar to
```

```
                transcriptional regulators e.g. glycine cleavage system
                transcription activator (gcv operon activator) -
                Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                identity 31 in 300 aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transcription regulator"
                /protein_id="BAB34178.1"
                /db_xref="GI:13360214"
                /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                EKSDSPHLQQFREWLLAKSVVPQEM"
gene            complement(3720. .4322)
                /gene="ECs0756"
CDS             complement(3720. .4322)
                /gene="ECs0756"
                /note="probable cob(I)alamin adenosyltransferase,similar
                to cob(I)alamin adenosyltransferases (corrinoid
                adenosyltransferases) e.g. [Escherichia coli]
                gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative cob(I)alamin adenosyltransferase"
                /protein_id="BAB34179.1"
                /db_xref="GI:13360215"
                /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene            complement(4332. .5984)
                /gene="ECs0757"
CDS             complement(4332. .5984)
                /gene="ECs0757"
                /note="probable fumarate hydratase,similar to fumarate
                hydratases e.g. fumarate hydratase class I, aerobic
                (fumarase) - Escherichia coli
                gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative fumarate hydratase"
                /protein_id="BAB34180.1"
                /db_xref="GI:13360216"
                /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSICGAAALLAQEYVKSLRCLEYPELG
                MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene            complement(6092. .7372)
                /gene="ECs0758"
CDS             complement(6092. .7372)
                /gene="ECs0758"
                /note="probable transport protein,similar to
                glutamate/aspartate transport proteins (proton glutamate
                symport proteins) e.g. [Bacillus stearothermophilus]
                gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                aa,also similar to C4-dicarboxylate transpor"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transport protein"
                /protein_id="BAB34181.1"
                /db_xref="GI:13360217"
                /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene            complement(7533. .7850)
                /gene="ECs0759"
CDS             complement(7533. .7850)
                /gene="ECs0759"
                /note="unknown"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="hypothetical protein"
                /protein_id="BAB34182.1"
```

Query Match 100.0%; Score 25; DB 1; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db   92367 GTTGCTGGATCTGCTGTTTGAAGCG 92391
```

```
RESULT  16
E00893
LOCUS       E00893       1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN
```

Query Match 96.0%; Score 24; DB 6; Length 1121;

```
  Best Local Similarity   100.0%;  Pred. No. 0.0055;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagc 24
           |||||||||||||||||||||||
Db      98 GTTGCTGGATCTGCTGTTTGAAGC 121


RESULT  17
AE004192
LOCUS       AE004192    12891 bp    DNA             BCT       31-JUL-2000
DEFINITION  Vibrio cholerae chromosome I, section 100 of 251 of the complete
            chromosome.
ACCESSION   AE004192 AE003852
VERSION     AE004192.1  GI:9655581
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 12891)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C. and White,O.
  TITLE     DNA sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
  JOURNAL   Nature 406 (6795), 477-483 (2000)
  MEDLINE   20406833
REFERENCE   2  (bases 1 to 12891)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12891
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /serotype="O1"
                     /db_xref="taxon:666"
                     /chromosome="I"
                     /note="biotype: El Tor"
     gene            complement(84. .1370)
                     /gene="VC1111"
     CDS             complement(84. .1370)
                     /gene="VC1111"
                     /note="similar to GB:J04423 SP:P12995 PID:457106
                     PID:490217 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="adenosylmethionine-8-amino-7-oxononanoate
                     aminotransferase"
                     /protein_id="AAF94270.1"
                     /db_xref="GI:9655582"
                     /translation="MMMDLAFDRCHIWHPYTSTLTPLTCYPVSRADGVMITLEDGRSL
                     VDGMSSWWAAIHGYNHPKLNQAAHQQIEQMSHIMFGGLTHQPAIELCQKLLKLAPNNL
                     QQVFLADSGSVAVEVSLKMALQYWHSKGQPRAKFLTLRHGYHGDTFAAMSVTDPDNSM
                     HSLYKGFLPEHIFANSPEGGFFDAWDERDIADFRHKLTEHHHQIAAVILEPIVQGAGG
                     MRIYHPEFLRQVRALCDEFGVLLILDEIATGFGRTGKLFACEHAGIQPDILCVGKALT
                     GGYMTLSATLTTQQVADTVCAGEAGCFMHGPTFMGNPLACAVASASLSLIEQGDWQQQ
                     VANIEAFFAEQLPKLNDSPRVKQTRWLGAIGVVETHLPVNMEVIQALFVEHGVWIRPF
                     GRLIYLMPPYISQPEHLQKLVDAMAAALNQPDCFKQ"
     gene            1497. .2549
                     /gene="VC1112"
     CDS             1497. .2549
                     /gene="VC1112"
                     /note="similar to GB:J04423 SP:P12996 PID:145425
                     PID:490219 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase"
                     /protein_id="AAF94271.1"
                     /db_xref="GI:9655583"
                     /translation="MEVRHNWTVAEVKALLDKPFMDLLFEAQQVHRLHHPHNHVQVST
                     LLSIKTGACPEDCKYCPQSAHYRTDVDKERLMEVERVLDAAQKAKNSGSTRFCMGAAW
                     KNPKERDMPLLKEMIRGVKDMGLETCMTLGMLTPDQAQQLAQAGLDYYNHNLDTSPEF
                     YGNIITTRTYQDRLDTLSHVRDAGMKICSGGIIGMGESTNDRAGLLVELANLPTHPES
                     VPINMLVKVKGTPLEQVDDVEPFDFVRLIAVARIMMPKSAVRLSAGREKMNEQMQALC
                     FMAGANSIFYGCKLLTTPNPAEDSDMLLFKKLGINREQVAQKPDEITENELLDRVVER
                     VAARPTASDLFYDAAL"
     gene            2533. .3687
                     /gene="VC1113"
     CDS             2533. .3687
                     /gene="VC1113"
                     /note="similar to GB:J04423 SP:P12998 PID:145426
                     PID:490225 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="8-amino-7-oxononanoate synthase"
                     /protein_id="AAF94272.1"
                     /db_xref="GI:9655584"
                     /translation="MTLRFKSRIEQALNEREHSRLTRRLQLLDRSAQGELIQQDRTYL
                     NFSSNDYLGLANDAELVQAWQTGLARFGAGSGASPMVTGFSSAHAELEHELCHWLGYE
                     RAVLFSSGFSANQALLFTLLEKEDLLLQDKLNHASLMEAGMLSPATMKRFKHNDTEHL
                     RQLLHEQSNSLVVTEGVFSMDGDCAPLAQIATLVKQRAWLMVDDAHGIGVLGEDGAGS
                     CQAAGIHPEILVVTFGKAFGLAGAAVLCDAHVGDYLTQFARHHVYSTAMPPAQAHALT
                     HAARMIQSQSWRREQLAELLACFDEQCRNIPGFVATQTPIKPWLLGSSDSALSASHEL
                     KQQGIWVSAIRPPTVPVGSARLRITITAAHTQAQIRRLSEQLKHVMEKDT"
     gene            3549. .4487
                     /gene="VC1114"
     CDS             3549. .4487
                     /gene="VC1114"
                     /note="similar to GB:J04423 SP:P12999 PID:145427
                     PID:490221 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthesis protein BioC"
                     /protein_id="AAF94273.1"
                     /db_xref="GI:9655585"
                     /translation="MGECHSTTHRACRQCAFTYHDNGGAYPSADTPLIGAVKTCDGEG
                     HMSMTATELCVELKDKSAIAQAFGKAAAHYDQHAAFQRDVGLRLLQKMPSCLKGLRVL
                     DLGCGTGYFSALLRERGAQVVCADISHAMLEQAKQRCGDEGMSYQLADAEQLPFASAC
                     FDMVFSSLALQWCEDLSLPLSEIRRVLKPHGQAFLSTLLDGSLFELEQAWRSVDHHRH
                     INQFISINQVKIALAQAGCSQHHLDLAAITVWYEAAFMLMRDLKGIGANHVSGRSTGL
                     ISRRTLAKVELAYQSFRNQQGLVPATYQVCLGVIHR"
     gene            4463. .5179
                     /gene="VC1115"
     CDS             4463. .5179
                     /gene="VC1115"
                     /note="similar to GB:S68059 SP:P13000 PID:490223
                     PID:145428 PID:460901; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAF94274.1"
                     /db_xref="GI:9655586"
                     /translation="MFRGYSPMRNAIFIAGTDTDVGKTVASKAILQALATHNIATIGY
                     KPVAAGSDKTEFGYRNSDALHLMKAATVDMPYEDVNPYALVLPTSPHIAAKHENVTID
                     YALLSNKLSKHKQNAELVVVEGAGGWRVPTSDSDCLSTWVKQERLPVILVVGIKLGCL
                     SHAILTAEAIRADGLELVGWIANRINPGTEHYAEIIEHLEGRLGTPKLGEIPYMPKAK
                     RQELGKFIQLDHLLEPDTVA"
     gene            complement(5261. .5470)
                     /gene="VC1116"
     CDS             complement(5261. .5470)
```

```
                     /gene="VC1116"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF94275.1"
                     /db_xref="GI:9655587"
                     /translation="MSRRNLTTNKGKHMAEAHSYQMAIDLLCCHLGISEQEARTQLGI.
                     EAKEQTLQQQIVDTQHSLMGLSAEK"
     gene            5697. .6560
                     /gene="VC1117"
     CDS             5697. .6560
                     /gene="VC1117"
                     /note="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
                     PID:1736470; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="heat shock protein HtpX"
                     /protein_id="AAF94276.1"
                     /db_xref="GI:9655588"
                     /translation="MKRILLFLATNLAVVLVLSVVLNIVYAVTGMQPGSLSGLLVMAA
                     VFGFGGAFISLLMSKSMALRSVGGVVIDTPRNEMEHWLLETVRRQANQAGIGMPTVAI
                     YDAPDMNAFATGAKRDDSLVAVSTGLLHNMTRDEAEAVLAHEVSHIANGDMVTMTLMQ
                     GVVNTFVIFLSRFIANIVASRDSEEGEGSNMMVYFGVSMVLELVFGFLASFITMWYSR
                     HREFHADAGAAQLVGKHKMIAALERLKMGQESHLEGSMMAFGITGKRSLSELMMTHPP
                     LEKRIAALRNM"
     gene            6770. .7198
                     /gene="VC1118"
     CDS             6770. .7198
                     /gene="VC1118"
                     /note="similar to PID:1653608; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, putative"
                     /protein_id="AAF94277.1"
                     /db_xref="GI:9655589"
                     /translation="MSIMDVQHVAQFYQQLDKSQLHRLIEIYHPDVVFEDAAHRIEGF
                     DALYQYFLNLYQNVHTCTFTIHEQYAVNEGAFLVWTMHLRHPKLAKGEQVDVKGVSHL
                     HFAEGKVTYHRDYFDMGEMLYEQLPVLGQVIRAIKRRLGQ"
     gene            7195. .7920
                     /gene="VC1119"
     CDS             7195. .7920
                     /gene="VC1119"
                     /note="similar to SP:P25970; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="oxidoreductase, short-chain
                     dehydrogenase/reductase family"
                     /protein_id="AAF94278.1"
                     /db_xref="GI:9655590"
                     /translation="MTAVFITGATSGIGKQLALDYAKQGWQVIACGRNQSVLDSLHTQ
                     YANIFPLAFDVTDHPGTKAALAQLPCQPELWILNAGDCEYIDDGKMDVTLMARVFNIN
                     VLGVAAVIEGIQPHLSCGHRVAIVGSIASELALPRAEAYGASKAAVAYLARTLQLDWR
                     PLGIEVTTIFPGFVATPLTDRNTFAMPMIITVERAAQEIKAGLARGVSQLYFPKRFTW
                     LIRLLGALPYAWQGRLVRRLLKA"
     gene            7931. .9211
                     /gene="VC1120"
     CDS             7931. .9211
                     /gene="VC1120"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11

  Query Match             80.0%;  Score 20;  DB 1;  Length 12891;
  Best Local Similarity   100.0%;  Pred. No. 0.68;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tggatctgctgtttgaagcg 25
           ||||||||||||||||||||
Db    1558 TGGATCTGCTGTTTGAAGCG 1577
```

```
RESULT  18
S64534/c
LOCUS       S64534        383 bp    DNA             INV       27-APR-2000
DEFINITION  Drosophila melanogaster defective chorion-1 eggshell protein
            (dec-1) gene, dec-1-Fc4 allele, partial cds.
ACCESSION   S64534
VERSION     S64534.1  GI:407575
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 383)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. II.
            Intraspecific DNA sequence analysis reveals length mutations in a
            repetitive region in D. melanogaster
  JOURNAL   J. Mol. Evol. 36 (6), 536-544 (1993)
  MEDLINE   93353525
   PUBMED   8350348
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 136269] from the original journal article.
            This sequence comes from Fig. 4.
            Map location: X 7C3-4.
FEATURES             Location/Qualifiers
     source          1. .383
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            <1. .>383
                     /gene="dec-1"
                     /allele="Fc4"
     CDS             <1. .>383
                     /gene="dec-1"
                     /note="This sequence comes from Fig. 5; follicle cell
                     protein for proper eggshell assembly; conceptual
                     translation presented here differs from translation in
                     publication"
                     /codon_start=1
                     /product="defective chorion-1 eggshell protein"
                     /protein_id="AAB27805.2"
                     /db_xref="GI:7548373"
                     /translation="QNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKI
                     QQNQQQIQQNPMMMQQRQWSEEQAKIQHDHQMAQQMAQQGLMMTEQRQRQWSEDQAKI
                     QQAQQMAQQTPMMMPQMQQRQWTEDP"
BASE COUNT      143 a     90 c    102 g     48 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgtt 18
           |||||||||||||||||
Db     159 TTGCTGGATCTGCTGTT 143


RESULT  19
S64532/c
LOCUS       S64532        461 bp    DNA             INV       27-APR-2000
DEFINITION  Drosophila melanogaster defective chorion-1 eggshell protein
            (dec-1) gene, dec-1-Fc3 allele, partial cds.
ACCESSION   S64532
VERSION     S64532.1  GI:407573
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 461)
```

```
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. II.
            Intraspecific DNA sequence analysis reveals length mutations in a
            repetitive region in D. melanogaster
  JOURNAL   J. Mol. Evol. 36 (6), 536-544 (1993)
  MEDLINE   93353525
   PUBMED   8350348
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 136268] from the original journal article.
            This sequence comes from Fig. 4.
            Map location: X 7C3-4.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            <1. .>461
                     /gene="dec-1"
                     /allele="Fc3"
     CDS             <1. .>461
                     /gene="dec-1"
                     /note="This sequence comes from Fig. 5; follicle cell
                     protein for proper eggshell assembly; conceptual
                     translation presented here differs from translation in
                     publication"
                     /codon_start=1
                     /product="defective chorion-1 eggshell protein"
                     /protein_id="AAB27804.2"
                     /db_xref="GI:7537267"
                     /translation="QNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKI
                     QQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMVMQQRQWSEEQAKIQHDQQM
                     AQQMAQQGLMMTEQRQRQWSEDQAK1QQAQQMAQQTPMMMPQMHQRQWTEDP"
BASE COUNT      174 a    108 c    122 g     57 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgtt 18
           |||||||||||||||||
Db     237 TTGCTGGATCTGCTGTT 221


RESULT  20
DMDECG1B/c
LOCUS       DMDECG1B      516 bp    DNA             INV       02-DEC-1993
DEFINITION  D.melanogaster dec-1 gene 3'.
ACCESSION   X76231
VERSION     X76231.1  GI:433551
KEYWORDS    dec-1 gene.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. I. Restriction
            site mapping and limited sequence comparison in the melanogaster
            species subgroup
  JOURNAL   J. Mol. Evol. 33 (4), 321-331 (1991)
  MEDLINE   92130257
COMMENT     X76230, X76231, X76232, x76233 are related accession numbers.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="7C3-4"
     gene            1. .516
                     /gene="dec-1"
     exon            <1. .>516
                     /gene="dec-1"
                     /number=2
     CDS             <1. .>516
                     /gene="dec-1"
                     /codon_start=1
                     /product="fc proteins"
                     /protein_id="CAA53816.1"
                     /db_xref="GI:433552"
                     /db_xref="FLYBASE:FBgn0000427"
                     /db_xref="SWISS-PROT:P18170"
                     /translation="KEAQRRARNSGiNTQKANALKRQAKSQDQTLSKEDIVQIMAYAY
                     RMANEQMESEKGKQDKVYAAYRTEQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQ
                     RQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEE
                     QAKIQQNQQQIQ"
     repeat_unit     205. .282
                     /gene="dec-1"
     repeat_unit     283. .360
                     /gene="dec-1"
     repeat_unit     361. .438
                     /gene="dec-1"
     repeat_unit     439. .516
                     /gene="dec-1"
BASE COUNT      194 a    120 c    137 g     65 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 516;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgtt 18
           |||||||||||||||||
Db     441 TTGCTGGATCTGCTGTT 425


RESULT  21
S64530/c
LOCUS       S64530        539 bp    DNA             INV       27-APR-2000
DEFINITION  Drosophila melanogaster defective chorion-1 eggshell protein
            (dec-1) gene, dec-1-Fc2 allele, partial cds.
ACCESSION   S64530
VERSION     S64530.1  GI:407569
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. II.
            Intraspecific DNA sequence analysis reveals length mutations in a
            repetitive region in D. melanogaster
  JOURNAL   J. Mol. Evol. 36 (6), 536-544 (1993)
  MEDLINE   93353525
   PUBMED   8350348
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 136266] from the original journal article.
            This sequence comes from Fig. 4.
            Map location: X 7C3-4.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /note="variant Israel"
     gene            <1. .>539
                     /gene="dec-1"
                     /allele="Fc2"
     CDS             <1. .>539
                     /gene="dec-1"
                     /note="This sequence comes from Fig. 5; follicle cell
                     protein for proper eggshell assembly; conceptual
                     translation presented here differs from translation in
```

```
                     publication"
                     /codon_start=1
                     /product="defective chorion-1 eggshell protein"
                     /protein_id="AAB27802.2"
                     /db_xref="GI:7537265"
                     /translation="QNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKI
                     QQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQQQ
                     IQQNPMMMQQRQWSEEQAKIQHDQQMAQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQM
                     AQQTPMMMPQMQQRQWTEDP"
BASE COUNT       208 a    124 c    141 g     66 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 539;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgtt 18
           ||||||||||||||||||
Db     315 TTGCTGGATCTGCTGTT 299


RESULT  22
S64531/c
LOCUS       S64531        539 bp    DNA             INV       27-APR-2000
DEFINITION  Drosophila melanogaster defective chorion-1 eggshell protein
            (dec-1) gene, dec-1-Fc2 allele, partial cds.
ACCESSION   S64531
VERSION     S64531.1  GI:407571
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. II.
            Intraspecific DNA sequence analysis reveals length mutations in a
            repetitive region in D. melanogaster
  JOURNAL   J. Mol. Evol. 36 (6), 536-544 (1993)
  MEDLINE   93353525
   PUBMED   8350348
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 136267] from the original journal article.
            This sequence comes from Fig. 4.
            Map location: X 7C3-4.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            <1. .>539
                     /gene="dec-1"
                     /allele="Fc2"
     CDS             <1. .>539
                     /gene="dec-1"
                     /note="This sequence comes from Fig. 5; follicle cell
                     protein for proper eggshell assembly; conceptual
                     translation presented here differs from translation in
                     publication"
                     /codon_start=1
                     /product="defective chorion-1 eggshell protein"
                     /protein_id="AAB27803.2"
                     /db_xref="GI:7537266"
                     /translation="QNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKI
                     QQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQ
                     IQQNPMMMQQRQWSEEQAKIQHDQQMAQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQM
                     AQQTPMMMPQMQQRQWTEDP"
BASE COUNT       209 a    124 c    140 g     66 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 539;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgtt 18
           ||||||||||||||||||
Db     315 TTGCTGGATCTGCTGTT 299


RESULT  23
S64529/c
LOCUS       S64529        617 bp    DNA             INV       27-APR-2000
DEFINITION  Drosophila melanogaster defective chorion-1 eggshell protein
            (dec-1) gene, dec-1-Fc1 allele, partial cds.
ACCESSION   S64529
VERSION     S64529.1  GI:407567
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Andersson,S. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. II.
            Intraspecific DNA sequence analysis reveals length mutations in a
            repetitive region in D. melanogaster
  JOURNAL   J. Mol. Evol. 36 (6), 536-544 (1993)
  MEDLINE   93353525
   PUBMED   8350348
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 136265] from the original journal article.
            This sequence comes from Fig. 4.
            Map location: X 7C3-4.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            <1. .>617
                     /gene="dec-1"
                     /allele="Fc1"
     CDS             <1. .>617
                     /gene="dec-1"
                     /note="This sequence comes from Fig. 5; follicle cell
                     protein for proper eggshell assembly; conceptual
                     translation presented here differs from translation in
                     publication"
                     /codon_start=1
                     /product="defective chorion-1 eggshell protein"
                     /protein_id="AAB27801.2"
                     /db_xref="GI:7537264"
                     /translation="QNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKI
                     QQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQ
                     IQQNPMMVQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQHDQQMAQQMAQ
                     QGLMMTEQRQRQWSEDQAKIQQAQQMAQQTPMMMPQMQQRQWTEDP"
BASE COUNT       239 a    143 c    160 g     75 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 617;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgtt 18
           ||||||||||||||||||
Db     393 TTGCTGGATCTGCTGTT 377


RESULT  24
DSDEC1TR6/c
LOCUS       DSDEC1TR6     774 bp    DNA             INV       25-SEP-1996
DEFINITION  D.sechellia DNA from repeated region of dec-1 locus.
ACCESSION   X96931
VERSION     X96931.1  GI:1262763
```

```
KEYWORDS    dec-1 locus; tandem repeat.
SOURCE      Drosophila sechellia.
  ORGANISM  Drosophila sechellia
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   Escher,S.A. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
  JOURNAL   Hereditas 124 (2), 111-120 (1996)
  MEDLINE   96376325
REFERENCE   2  (bases 1 to 774)
  AUTHORS   Escher,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
            Genetics, S-901 87 Umea, SWEDEN
COMMENT     Related sequences: M35889, M35688 and X07697.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Drosophila sechellia"
                     /db_xref="taxon:7238"
                     /chromosome="X"
                     /map="7C"
     repeat_region   1. .774
                     /rpt_type=TANDEM
     gene            1. .774
                     /gene="dec-1"
     exon            <1. .>774
                     /gene="dec-1"
                     /number=2
BASE COUNT      293 a    178 c    213 g     90 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 774;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||
Db     81 TTGCTGGATCTGCTGTT 65


RESULT  25
DSDEC1TR2/c
LOCUS       DSDEC1TR2     852 bp    DNA             INV       25-SEP-1996
DEFINITION  D.simulans DNA from repeated region of dec-1 locus, strain
            Yugoslavia.
ACCESSION   X96927
VERSION     X96927.1  GI:1262760
KEYWORDS    dec-1 locus; tandem repeat.
SOURCE      Drosophila simulans.
  ORGANISM  Drosophila simulans
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   Escher,S.A. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
  JOURNAL   Hereditas 124 (2), 111-120 (1996)
  MEDLINE   96376325
REFERENCE   2  (bases 1 to 852)
  AUTHORS   Escher,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
            Genetics, S-901 87 Umea, SWEDEN
COMMENT     Related sequences: M35889, M35688 and X07697.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Drosophila simulans"
                     /strain="Yugoslavia"
                     /db_xref="taxon:7240"
                     /chromosome="X"
                     /map="7C"
     repeat_region   1. .852
                     /rpt_type=TANDEM
     gene            1. .852
                     /gene="dec-1"
     exon            <1. .>852
                     /gene="dec-1"
                     /number=2
BASE COUNT      324 a    195 c    232 g    101 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 852;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||
Db    237 TTGCTGGATCTGCTGTT 221


RESULT  26
DSDEC1TR3/c
LOCUS       DSDEC1TR3     852 bp    DNA             INV       25-SEP-1996
DEFINITION  D.simulans DNA from repeated region of dec-1 locus, strain
            Barcelona.
ACCESSION   X96928
VERSION     X96928.1  GI:1262761
KEYWORDS    dec-1 locus; tandem repeat.
SOURCE      Drosophila simulans.
  ORGANISM  Drosophila simulans
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   Escher,S.A. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
  JOURNAL   Hereditas 124 (2), 111-120 (1996)
  MEDLINE   96376325
REFERENCE   2  (bases 1 to 852)
  AUTHORS   Escher,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
            Genetics, S-901 87 Umea, SWEDEN
COMMENT     Related sequences: M35889, M35688 and X07697.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Drosophila simulans"
                     /strain="Barcelona"
                     /db_xref="taxon:7240"
                     /chromosome="X"
                     /map="7C"
     repeat_region   1. .852
                     /rpt_type=TANDEM
     gene            1. .852
                     /gene="dec-1"
     exon            <1. .>852
                     /gene="dec-1"
                     /number=2
BASE COUNT      324 a    195 c    232 g    101 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 852;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||
Db    237 TTGCTGGATCTGCTGTT 221


RESULT  27
DMDEC1TR1/c
LOCUS       DMDEC1TR1     930 bp    DNA             INV       25-SEP-1996
DEFINITION  D.melanogaster DNA from repeated region of dec-1 locus, strain
            Samarkand.
ACCESSION   X96926
VERSION     X96926.1  GI:1262758
KEYWORDS    dec-1 locus; tandem repeat.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 930)
  AUTHORS   Escher,S.A. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
  JOURNAL   Hereditas 124 (2), 111-120 (1996)
  MEDLINE   96376325
REFERENCE   2  (bases 1 to 930)
  AUTHORS   Escher,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
            Genetics, S-901 87 Umea, SWEDEN
COMMENT     Related sequences: M35889, M35688 and X07697.
FEATURES             Location/Qualifiers
     source          1. .930
                     /organism="Drosophila melanogaster"
                     /strain="Samarkand"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="7C"
     repeat_region   1. .930
                     /rpt_type=TANDEM
     gene            1. .930
                     /gene="dec-1"
     exon            <1. .>930
                     /gene="dec-1"
                     /number=2
BASE COUNT      352 a    213 c    254 g    111 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 930;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||
Db    393 TTGCTGGATCTGCTGTT 377


RESULT  28
DRODEC1A/c
LOCUS       DRODEC1A     3741 bp    mRNA            INV       28-NOV-2000
DEFINITION  Drosophila melanogaster defective chorion-1 fcl25 protein precursor
            (dec-1) mRNA, complete cds.
ACCESSION   M35887 M35688
VERSION     M35887.2  GI:11386620
KEYWORDS    dec-1 protein; eggshell protein.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78)
  AUTHORS   Hawley,R.J. and Waring,G.L.
  TITLE     Cloning and analysis of the dec-1 female-sterile locus, a gene
            required for proper assembly of the Drosophila eggshell
  JOURNAL   Genes Dev. 2 (3), 341-349 (1988)
  MEDLINE   88243015
   PUBMED   3378704
REFERENCE   2  (bases 1 to 3741)
  AUTHORS   Waring,G.L., Hawley,R.J. and Schoenfeld,T.
  TITLE     Multiple proteins are produced from the dec-1 eggshell gene in
            Drosophila by alternative RNA splicing and proteolytic cleavage
            events
  JOURNAL   Dev. Biol. 142 (1), 1-12 (1990)
  MEDLINE   91032553
   PUBMED   1699826
COMMENT     On Nov 28, 2000 this sequence version replaced gi:157181.
            Draft entry and computer-readable sequence kindly submitted by
            G.L.Waring, 28-JUN-1990.
            Sequence update by GenBank staff to remove sequence derived from
            genomic DNA. which can be found in X07697.
FEATURES             Location/Qualifiers
     source          1. .3741
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="7C"
                     /tissue_type="ovary"
     gene            1. .3741
                     /gene="dec-1"
     CDS             76. .3447
                     /gene="dec-1"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="defective chorion-1 fcl25 protein precursor"
                     /protein_id="AAA28446.1"
                     /db_xref="GI:157182"
                     /translation="MRLFSLLPLLALLVVQVAGQSEVTSDDPATDAGSTTNSTTDTKP
                     RIPSQDEILGQMPSINPIRTGNPQMDAFYMMFPALGSLLKWGSLFPAYSILGAIPDNL
                     QPTAAASKVVLVLADDATAKTRVARQNPPPNPLGQLMNWPALPQDFQLPSMDLGPQVG
                     SFLAQLPAMPTVPGLLGAAAPVPAPAPAPAAAPPPAPAPAADPPAAPVPDAPQPAILG
                     EAALQNAFTFFNPANFDASSLLGQSVPTFAPPNLDFVAQMQRQFFPGMTPAQPAAAGT
                     DAQASDISEVRVRPEDPYSQEAQMKIKSALEMEQERQQQAQVKDQEQVPLLWFRMPTT
                     QNQDATEEKTLEHLRVEAKLRAFERQVIAELRMLQKIELMAKQMRSSTAAQNGDSPYR
                     ISYPLSRTPIHKITRADIEQALRDDYVRRLVNKEAQRRARNSGINTQKANALKRQAKS
                     QDQTLSKEDIVQIMAYAYRMANEQMESEKGKQDKVYAAYRTEQNPMMMQQRQWSEEQA
                     KIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQ
                     QQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQQQIQQN
                     PMMMQQRQWSEEQAKIQHDQQMAQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQMAQQT
                     PMMMPQMQQRQWTEDPQMVQQMQQRQWAEDQTRMQMAQQNPMMQQQRQMAENPQMMQQ
                     RQWSEEQTKIEQAQQMAQQNQMMMQQMQQRQWSEDQAQIQQQQRQMMQQTPMMMKERQ
                     WAEENPQSVQQQGPMMMQQQMPSMMQREVEDEDNKAEDDLVGEAGPQMAGERRHRQAQ
                     SRCPGSWRQQAQEVQVQVGAATVINYYYAAPSVRWFRVTEQATAEGGYGSNAYGVPRP
                     VNSYQSQGYRAAVGNDEVDEMLRQHQTMARTINPKQPGEVGGSESQKSNSNPPTTLTP
                     APQEQPQEHRVHKSPSSAPSETEIENAPSSDPQVGSIFTYGEGLLHPFMGLLPVERPD
                     DPWNQKPYDPHHPLYTGGGSYDAYLRDGRHRRDTHIMGQGTQHGILTPGMLERLLRIK
                     MDFQRRFPPFIQGYAESPYESYPCRGTASGSGQDF"
     sig_peptide     76. .144
                     /gene="dec-1"
     mat_peptide     145. .3444
                     /gene="dec-1"
                     /product="defective chorion-1 fcl25 protein"
BASE COUNT     1091 a    987 c    993 g    670 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 3741;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||
Db   1944 TTGCTGGATCTGCTGTT 1928


RESULT  29
```

```
DRODEC1B/c
LOCUS       DRODEC1B     3755 bp    mRNA            INV       28-NOV-2000
DEFINITION  Drosophila melanogaster defective chorion-1 fc106 protein precursor
            (dec-1) mRNA, complete cds.
ACCESSION   M35888 M35688
VERSION     M35888.2  GI:11386621
KEYWORDS    dec-1 protein; eggshell protein.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78)
  AUTHORS   Hawley,R.J. and Waring,G.L.
  TITLE     Cloning and analysis of the dec-1 female-sterile locus, a gene
            required for proper assembly of the Drosophila eggshell
  JOURNAL   Genes Dev. 2 (3), 341-349 (1988)
  MEDLINE   88243015
   PUBMED   3378704
REFERENCE   2  (bases 1 to 3755)
  AUTHORS   Waring,G.L., Hawley,R.J. and Schoenfeld,T.
  TITLE     Multiple proteins are produced from the dec-1 eggshell gene in
            Drosophila by alternative RNA splicing and proteolytic cleavage
            events
  JOURNAL   Dev. Biol. 142 (1), 1-12 (1990)
  MEDLINE   91032553
   PUBMED   1699826
COMMENT     On Nov 28, 2000 this sequence version replaced gi:157183.
            Draft entry and computer-readable sequence kindly submitted by
            G.L.Waring, 28-JUN-1990.
            Sequence update by GenBank staff to remove sequence derived from
            genomic DNA. which can be found in X07697.
FEATURES             Location/Qualifiers
     source          1. .3755
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="7C"
                     /tissue_type="ovary"
     gene            1. .3755
                     /gene="dec-1"
     CDS             76. .2928
                     /gene="dec-1"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="defective chorion-1 fc106 protein precursor"
                     /protein_id="AAA28447.1"
                     /db_xref="GI:157184"
                     /translation="MRLFSLLPLLALLVVQVAGQSEVTSDDPATDAGSTTNSTTDTKP
                     RIPSQDEILGQMPSINPIRTGNPQMDAFYMMFPALGSLLKWGSLFPAYSILGAIPDNL
                     QPTAAASKVVLVLADDATAKTRVARQNPPPNPLGQLMNWPALPQDFQLPSMDLGPQVG
                     SFLAQLPAMPTVPGLLGAAAPVPAPAPAPAAAPPPAPAPAADPPAAPVPDAPQPAILG
                     EAALQNAFTFFNPANFDASSLLGQSVPTFAPPNLDFVAQMQRQFFPGMTPAQPAAAGT
                     DAQASDISEVRVRPEDPYSQEAQMKIKSALEMEQERQQQAQVKDQEQVPLLWFRMPTT
                     QNQDATEEKTLEHLRVEAKLRAFERQVIAELRMLQKIELMAKQMRSSTAAQNGDSPYR
                     ISYPLSRTPIHKITRADIEQALRDDYVRRLVNKEAQRRARNSGINTQKANALKRQAKS
                     QDQTLSKEDIVQIMAYAYRMANEQMESEKGKQDKVYAAYRTEQNPMMMQQRQWSEEQA
                     KIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQ
                     QQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQQQIQQN
                     PMMMQQRQWSEEQAKIQHDQQMAQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQMAQQT
                     PMMMPQMQQRQWTEDPQMVQQMQQRQWAEDQTRMQMAQQNPMMQQQRQMAENPQMMQQ
                     RQWSEEQTKIEQAQQMAQQNQMMMQQMQQRQWSEDQAQIQQQQRQMMQQTPMMMKERQ
                     WAEENPQSVQQQGPMMMQQQMPSMMQREVEDEDNKAEDDLVGEAGPQMAGERRHRQAQ
                     SRCPGSWRQQAQEVQVQVGAATVINYYYAAPSVRWFRVTEQATAEGGYGSNAYGVPRP
                     VNSYQSQGYRAAVGNDEVDEMLRQHQTMARATHFRQ"
     sig_peptide     76...144
                     /gene="dec-1"
     mat_peptide     145. .2925
                     /gene="dec-1"
                     /product="defective chorion-1 fc106 protein"
BASE COUNT     1094 a    991 c    997 g    673 t
ORIGIN
```

```
  Query Match           68.0%;  Score 17;  DB 3;  Length 3755;
  Best Local Similarity  100.0%;  Pred. No. 32;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgtt 18
           |||||||||||||||||
Db    1944 TTGCTGGATCTGCTGTT 1928
```

```
RESULT  30
DRODEC1C/c
LOCUS       DRODEC1C     5697 bp    mRNA            INV       28-NOV-2000
DEFINITION  Drosophila melanogaster defective chorion-1 fc177 protein precursor
            (dec-1) mRNA, complete cds.
ACCESSION   M35889 M35688
VERSION     M35889.2  GI:11386622
KEYWORDS    dec-1 protein; eggshell protein.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 78)
  AUTHORS   Hawley,R.J. and Waring,G.L.
  TITLE     Cloning and analysis of the dec-1 female-sterile locus, a gene
            required for proper assembly of the Drosophila eggshell
  JOURNAL   Genes Dev. 2 (3), 341-349 (1988)
  MEDLINE   88243015
   PUBMED   3378704
REFERENCE   2  (bases 1 to 5697)
  AUTHORS   Waring,G.L., Hawley,R.J. and Schoenfeld,T.
  TITLE     Multiple proteins are produced from the dec-1 eggshell gene in
            Drosophila by alternative RNA splicing and proteolytic cleavage
            events
  JOURNAL   Dev. Biol. 142 (1), 1-12 (1990)
  MEDLINE   91032553
   PUBMED   1699826
COMMENT     On Nov 28, 2000 this sequence version replaced gi:157185.
            Draft entry and computer-readable sequence kindly submitted by
            G.L.Waring, 28-JUN-1990.
            Sequence update by GenBank staff to remove sequence derived from
            genomic DNA. which can be found in X07697.
FEATURES             Location/Qualifiers
     source          1. .5697
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
                     /map="7C"
                     /tissue_type="ovary"
     gene            1. .5697
                     /gene="dec-1"
     CDS             76. .4845
                     /gene="dec-1"
                     /note="alternatively spliced"
                     /codon_start=1
                     /product="defective chorion-1 fc177 protein precursor"
                     /protein_id="AAA28448.1"
                     /db_xref="GI:157186"
                     /translation="MRLFSLLPLLALLVVQVAGQSEVTSDDPATDAGSTTNSTTDTKP
                     RIPSQDEILGQMPSINPIRTGNPQMDAFYMMFPALGSLLKWGSLFPAYSILGAIPDNL
                     QPTAAASKVVLVLADDATAKTRVARQNPPPNPLGQLMNWPALPQDFQLPSMDLGPQVG
                     SFLAQLPAMPTVPGLLGAAAPVPAPAPAPAAAPPPAPAPAADPPAAPVPDAPQPAILG
                     EAALQNAFTFFNPANFDASSLLGQSVPTFAPPNLDFVAQMQRQFFPGMTPAQPAAAGT
                     DAQASDISEVRVRPEDPYSQEAQMKIKSALEMEQERQQQAQVKDQEQVPLLWFRMPTT
                     QNQDATEEKTLEHLRVEAKLRAFERQVIAELRMLQKIELMAKQMRSSTAAQNGDSPYR
                     ISYPLSRTPIHKITRADIEQALRDDYVRRLVNKEAQRRARNSGINTQKANALKRQAKS
                     QDQTLSKEDIVQIMAYAYRMANEQMESEKGKQDKVYAAYRTEQNPMMMQQRQWSEEQA
                     KIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQ
                     QQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQQQIQQN
                     PMMMQQRQWSEEQAKIQHDQQMAQQMAQQGLMMTEQRQRQWSEDQAKIQQAQQMAQQT
                     PMMMPQMQQRQWTEDPQMVQQMQQRQWAEDQTRMQMAQQNPMMQQQRQMAENPQMMQQ
                     RQWSEEQTKIEQAQQMAQQNQMMMQQMQQRQWSEDQAQIQQQQRQMMQQTPMMMKERQ
```

```
                     WAEENPQSVQQQGPMMMQQQMPSMMQREVEDEDNKAEDDLVGEAGPQMAGERRHRQAQ
                     SRCPGSWRQQAQEVQVQVGAATVINYYYAAPSVRWFRVTEQATAEGGYGSNAYGVPRP
                     VNSYQSQGYRAAVGNDEVDEMLRQHQTMARTINPKQPGEVGGSESQKSNSNPPTTLTP
                     APQEQPQEHRVHKRLAHFHRFGQEAGLNATTSKGCGCGRLDCLCGRSCRCGRRGLESR
                     VVSSRTSGTCQCKASHRNKRCEYGTLETIDEGSLNELRREYKLGLKEITLSPDEDPAE
                     ALMRYNAASIREALERASMEPLEIGGDQYEEDAQQEPMEEEQLQHDPNTEPQYNHKDF
                     VRLTTSTASPITSTTEAATPTGSDSTSEATVTPEVTTTTSTSTTTTESTKDEGLDMQ
                     QDSQAEAESSHVTKSISKQEAXIHQLHSIVEELKNEILKLNLRCSTIISNNVAKEPVT
                     EKNPPVVEEPSKQEDKPKVEEKVIAEEQAPVEPEEDLEEDEDSTSISTTTETPSPSGS
                     YSTKPGLSLGSPRVDEQSGSSNKLDYDDDNNWQRILANRGYDTDYLTKSHERQFAQGQ
                     NLEMPKNCNYDGNGSQEYGPYPEFQADEPSTDTEGKAKRALSVKQQAQLLNAALNDSG
                     SDSSDGTTTTTTPSPYAMRGKFVRRRSTARRVPIPKIGKASDEVWVRSPRQAKMPQRP
                     KKSMSKPKKQSSQVTTQATVSSTKLDSLVDVLKDLVRLQIQKEKKSSLLRTQSNNLSK
                     TKPKSIKPVKVIKRKRLRRRQHKSIATTIRSPIQTKA"
     sig_peptide     76. .144
                     /gene="dec-1"
     mat_peptide     145. .4842
                     /gene="dec-1"
                     /product="defective chorion-1 fc177 protein"
BASE COUNT     1704 a   1469 c   1485 g   1038 t      1 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 3;  Length 5697;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgtt 18
           |||||||||||||||||
Db    1944 TTGCTGGATCTGCTGTT 1928


RESULT  31
AC017995/c
LOCUS       AC017995    33204 bp    DNA             HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017995
VERSION     AC017995.1  GI:6553195
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 33204)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212937 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .33204
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     8981 a   7720 c   7703 g   8800 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 33204;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgtt 18
           |||||||||||||||||
Db   15905 TTGCTGGATCTGCTGTT 15889


RESULT  32
AC016952/c
LOCUS       AC016952   161540 bp    DNA             HTG       27-AUG-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-220O14, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC016952
VERSION     AC016952.11  GI:9929556
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161540)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161540)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Aug 26, 2000 this sequence version replaced gi:9719634.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HMUN
                Center clone name: RP11-220O14
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 138744 bases at least Q40
                Consensus quality: 149435 bases at least Q30
                Consensus quality: 153457 bases at least Q20
                Estimated insert size: 155950; sum-of-contigs estimation
                Estimated insert size: 211269; agarose-fp estimation
                Quality coverage: 3.5x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5164: contig of 5164 bp in length
            *     5165     5264: gap of unknown length
```

```
            *     5265    55000: contig of 49736 bp in length
            *    55001    55100: gap of unknown length
            *    55101    84235: contig of 29135 bp in length
            *    84236    84335: gap of unknown length
            *    84336   104093: contig of 19758 bp in length
            *   104094   104193: gap of unknown length
            *   104194   117951: contig of 13758 bp in length
            *   117952   118051: gap of unknown length
            *   118052   126732: contig of 8681 bp in length
            *   126733   126832: gap of unknown length
            *   126833   135507: contig of 8675 bp in length
            *   135508   135607: gap of unknown length
            *   135608   141357: contig of 5750 bp in length
            *   141358   141457: gap of unknown length
            *   141458   145691: contig of 4234 bp in length
            *   145692   145791: gap of unknown length
            *   145792   149739: contig of 3948 bp in length
            *   149740   149839: gap of unknown length
            *   149840   152725: contig of 2886 bp in length
            *   152726   152825: gap of unknown length
            *   152826   155298: contig of 2473 bp in length
            *   155299   155398: gap of unknown length
            *   155399   158289: contig of 2891 bp in length
            *   158290   158389: gap of unknown length
            *   158390   160296: contig of 1907 bp in length
            *   160297   160396: gap of unknown length
            *   160397   161540: contig of 1144 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161540
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-220O14"
BASE COUNT    53135 a  28040 c  27770 g  51176 t   1419 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 2;  Length 161540;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctgttt 19
           |||||||||||||||||
Db   22839 TGCTGGATCTGCTGTTT 22823


RESULT  33
AE003442
LOCUS       AE003442   306267 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386054 section 26
            of 35, complete sequence.
ACCESSION   AE003442 AE002566
VERSION     AE003442.1  GI:7290831
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 306267)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe.M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 306267)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .306267
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
     mRNA            join(32686. .32823,32937. .33134,33225. .33387,34597. .34762,
                     35618. .35749)
                     /gene="CG1402"
                     /note="Partial transcript."
                     /product="CT3206"
                     /db_xref="FLYBASE:FBan0001402"
                     /db_xref="FLYBASE:FBgn0029962"
     gene            <32686. .>35749
                     /gene="CG1402"
                     /map="7C3-7C4"
                     /db_xref="FLYBASE:FBan0001402"
                     /db_xref="FLYBASE:FBgn0029962"
     CDS             join(33264. .33387,34597. .34762,35618. .35639)
                     /gene="CG1402"
                     /note="CG1402 gene product; Partial transcript"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0001402"
                     /db_xref="FLYBASE:FBgn0029962"
                     /protein_id="AAF46274.1"
                     /db_xref="GI:7290832"
                     /translation="MSEAQHKSQGIVGLSLMVQIGETPNPELRIITSTFNKVLYRGFS
                     TPIRHISVRSLLPNTDHYITYEGSTTHPGCWESTVWIIVNKPIYITKQEVSLCCTNSD
                     D"
     mRNA            complement(48924. .>50117)
                     /gene="CG10920"
                     /product="CT30589"
                     /db_xref="FLYBASE:FBan0010920"
                     /db_xref="FLYBASE:FBgn0029963"
     gene            complement(<48924. .>50117)
                     /gene="CG10920"
                     /map="7C3-7C5"
                     /db_xref="FLYBASE:FBan0010920"
                     /db_xref="FLYBASE:FBgn0029963"
     CDS             complement(48924. .50033)
                     /gene="CG10920"
```

```
                /note="CG10920 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0010920"
                /db_xref="FLYBASE:FBgn0029963"
                /protein_id="AAF46275.1"
                /db_xref="GI:7290833"
                /translation="MRHPTPPPLAVQHARSGAMVVELAVDPSRAMRSTNQQQPIDRDQ
                LRISTEESCQGSELVVASSHLSKKIQEHRRPNQLVRFCLRLGYNTLLYPYEMAKVLIQ
                LGHEPLQAKPFVMRLIQRRPRLFLPSVHRYVQHIQHIDGYTGMYRGLTARLAASIVDY
                LLGDLLLATLHFAPYKRGPKEGLSLKEFWWNLTRNSLRITTVVVITHPFYVVMVRQIA
                QFVGGEHVYEGLVGSLMTLAQQEGCAGLFAGMVPRLLGEWSVLFITSALSHLCRRLLP
                MSELQHQYNTAVIQMMASLVAYPLEVTSTCMAGTGAPLTACEPPSMPLYNHWVDCLTD
                LYARGGQNRGAILFWRTVPKIQLLRSKDNYTVRSI"
mRNA            <54229. .>54681
                /gene="CG1409"
                /product="CT3292"
                /db_xref="FLYBASE:FBan0001409"
                /db_xref="FLYBASE:FBgn0029964"
gene            <54229. .>54681
                /gene="CG1409"
                /map="7C4-7C5"
                /db_xref="FLYBASE:FBan0001409"
                /db_xref="FLYBASE:FBgn0029964"
CDS             54229. .54681
                /gene="CG1409"
                /note="CG1409 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0001409"
                /db_xref="FLYBASE:FBgn0029964"
                /protein_id="AAF46276.1"
                /db_xref="GI:7290834"
                /translation="MARYLAQIIILGAQLVGRALVKTMRQELQAFEDAARLQETLKAN
                DPNSGRSAVAKTMTLAEAQQILDVSDLTNRQAIDTHYQHLFRVNDKSTGGSFYIQSKV
                FRAKERIDQELERTELLVKTDDSHSILPTPPESSQFQCENKEPGQKSR"
mRNA            complement(join(<60589. .60716,60988. .62089,62143. .>62916))
                /gene="CG15326"
                /product="CT35326"
                /db_xref="FLYBASE:FBan0015326"
                /db_xref="FLYBASE:FBgn0029965"
                /evidence=not_experimental
gene            complement(<60589. .>62916)
                /gene="CG15326"
                /map="7C4-7C5"
                /db_xref="FLYBASE:FBan0015326"
                /db_xref="FLYBASE:FBgn0029965"
                /evidence=not_experimental
CDS             complement(join(60589. .60716,60988. .62089,62143. .62916))
                /gene="CG15326"
                /note="CG15326 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0015326"
                /db_xref="FLYBASE:FBgn0029965"
                /evidence=not_experimental
                /protein_id="AAF46277.1"
                /db_xref="GI:7290835"
                /translation="MESSSDWDLAEALAQVVANSEMGRFKTLYIYTHTNSQSTGGHLE
                ELLDQVLMIVPNNLQARRLLLQQSMEYKPYVHAVLALVDGLPSLSAIYARIRATQDLS
                HTLIYMSMPTDAYGEEMQATLRFLWRLSVLNVGVVLRPPGDHILMVSYFPFSALHGCQ
                VISANVVNRYQVGTKRWASQDYFPSKLGNFYGCLLTCATWEDMPYLVWRPDGSGSFVG
                IEGALLQFMAENLNFTVGLYWMNKEEVLATFDESGRIFDEIFGHHADFSLGGFHFKPS
                AGSEIPYSQSTYYFMSHIMLVTNLQSAYSAYEKLSFPFTPLIWRAIGLVLILACLLLM
                LLVRWRHHHELPRNPYYELLVLTMGGNLEDRWVPQRFPSRLVLLTWLFATLVLRSGYQ
                SGMYQLLRQDTQRNPPQTISEVLAQHFTIQLAEVNEARILASLPELRPEQLVYLEGSE
                LQSFPALAQQSGSSARVAILTPYEYFGYFRKVHPMSRRLHLVRERIYTQQLAFYVRRH
                SHLVGVLNKQIQHAHTHGFLEHWTRQYVSAVDEKDESVARIASTSYSTLDGIDGDPSL
                SESEEDQQVAPVRQNVLSMRELAALFWLILWANLGAVVVFVLELLLPRIKLRKILRKM
                KKSTRASATTTSTLSSPSTTKDIPFSCKDGFQDSWPKCSLLVS"
mRNA            complement(join(63233. .63902,64113. .65786,65862. .65986,
                66185. .66447,66505. .68920,68994. .>69146))
                /gene="dec-1"
                /note="Nucleotide sequence of the Celera sequence differs
                from the published sequence for this transcript."
                /product="CT7110"
                /db_xref="FLYBASE:FBan0002175"
                /db_xref="FLYBASE:FBgn0000427"
gene            complement(<63233. .>69146)
                /gene="dec-1"
                /note="CG2175"
                /map="7C4-7C5"
                /db_xref="FLYBASE:FBan0002175"
                /db_xref="FLYBASE:FBgn0000427"
CDS             complement(join(63233. .63902,64113. .65786,65862. .65986,
                66185. .66447,66505. .68854))
                /gene="dec-1"
                /note="dec-1 gene product; Nucleotide sequence of the
                Celera sequence differs from the published sequence for
                this transcript"
                /codon_start=1
                /db_xref="FLYBASE:FBan0002175"
                /db_xref="FLYBASE:FBgn0000427"
                /protein_id="AAF46278.1"
                /db_xref="GI:7290836"
                /translation="MMFPALGSLLKWGSLFPAYSILGAIPDNLQPTAAASKVVLVLAD
                DATAKTRVARQNPPPNPLGQLMNWPALPQDFQLPSMDLGPQVGSFLAQLPAMPTVPGL
                LGAAAPVPAPAPAPAAAPPPAPAPAADPPAAPVPDAPQPAILGQAALQNAFTFFNPAN
                FDASSLLGQSVPTFAPPNLDFVAQMQRQFFPGMTPAQPAAAGTDAQASDISEVRVRPE
                DPYSQEAQMKIKSALEMEQERQQQAQVKDQEQVPLLWFRMPTTQNQDATEEKTLEDLR
                VEAKLRAFERQVIAELRMLQKIELMAKQMRSSAAAQNGDSPYRISYPLSRTPIHKITR
                ADIEQALRDDYVRRLVNKEAQRRARNSGINTQKANALKRQAKSQDQTLSKEDIVQIMA
                YAYRMANEQMESEKGKQDKVYAAYRTEQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMM
                MQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQW
                SEEQAKIQQNQQQIQQNPMMVQQRQWSEEQAKIQQNQQQIQQNPMMMQQRQWSEEQAK
```

Query Match 68.0%; Score 17; DB 3; Length 306267;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      2 ttgctggatctgctgtt 18
          ||||||||||||||||
Db  67205 TTGCTGGATCTGCTGTT 67221
```

```
RESULT  34
DMDEC1TR5/c
LOCUS       DMDEC1TR5     774 bp    DNA             INV       25-SEP-1996
DEFINITION  D.mauritiana DNA from repeated region of dec-1 locus.
ACCESSION   X96930
VERSION     X96930.1  GI:1262759
KEYWORDS    dec-1 locus; tandem repeat.
SOURCE      Drosophila mauritiana.
  ORGANISM  Drosophila mauritiana
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 774)
  AUTHORS   Escher,S.A. and Lambertsson,A.
  TITLE     Evolution of the dec-1 eggshell locus in Drosophila. III. Sequence
            comparisons of the simulans complex repeated domain reveal
            non-concerted evolution
  JOURNAL   Hereditas 124 (2), 111-120 (1996)
  MEDLINE   96376325
REFERENCE   2  (bases 1 to 774)
  AUTHORS   Escher,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) S.A. Escher, Umea University, Department of
            Genetics, S-901 87 Umea, SWEDEN
COMMENT     Related sequences: M35889, M35688 and X07697.
FEATURES             Location/Qualifiers
     source          1. .774
                     /organism="Drosophila mauritiana"
                     /db_xref="taxon:7226"
                     /chromosome="X"
                     /map="7C"
     repeat_region   1. .774
                     /rpt_type=TANDEM
     gene            1. .774
```

```
                     /gene="dec-1"
     exon            <1. .>774
                     /gene="dec-1"
                     /number=2
BASE COUNT      290 a    182 c    212 g     90 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 3;  Length 774;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgctgt 17
          ||||||||||||||||
Db    159 TTGCTGGATCTGCTGT 144


RESULT  35
AF250347/c
LOCUS       AF250347     1044 bp    mRNA            VRT       01-JUL-2001
DEFINITION  Xenopus laevis transcription factor NKX-2.1 mRNA, complete cds.
ACCESSION   AF250347
VERSION     AF250347.1  GI:14578835
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1044)
  AUTHORS   Hollemann,T. and Pieler,T.
  TITLE     Xnkx-2.1: a homeobox gene expressed during early forebrain, lung
            and thyroid development in Xenopus laevis
  JOURNAL   Dev. Genes Evol. 210 (11), 579-581 (2000)
  MEDLINE   21110556
   PUBMED   11180810
REFERENCE   2  (bases 1 to 1044)
  AUTHORS   Hollemann,T. and Pieler,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2000) Entwicklungsbiochemie, University of
            Goettingen, Humboldtallee 23, Goettingen, LowerSaxony 37073,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1044
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     primer_bind     1. .29
     CDS             1. .1044
                     /note="Xnkx-2.1"
                     /codon_start=1
                     /product="transcription factor NKX-2.1"
                     /protein_id="AAK69025.1"
                     /db_xref="GI:14578836"
                     /translation="MSMSPKHTTPFSVSDILSPLEESYKKVAMEGAGLGAPLTAAAYR
                     QSQVSQASMQQHGMGHNGPVSAAYHMTAAGVPQLSHTTMGGYCNGNLGNLGNMSELPP
                     YQDTMRNSSATGWYGANPDPRFSTISRFMGPSGGMNMEGLGNMGSLGDVGKSMTPLQA
                     TPRRKRRVLFSQAQVYELERRFKQQKYLSAPEREHLASMIHLTPTQVKIWFQNHRYKM
                     KRQAKDKASQQQMQQDNSSCQQQQSPRRVAVPVLVKDGKPCQAGSNTPTAALQSHQQQ
                     TATAITITTNGLGPHQSHQTNSAGQSPDLVPHSNSPSSLQNQVTSLSHLNSSSSDYGS
                     AMSCSTLLYGRTW"
BASE COUNT      277 a    311 c    265 g    190 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 5;  Length 1044;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaagcg 25
           ||||||||||||||||
Db    556 TCTGCTGTTTGAAGCG 541
```

```
RESULT  36
AF281080/c
LOCUS       AF281080     1614 bp    mRNA            VRT       01-OCT-2000
DEFINITION  Xenopus laevis homeobox transcription factor Nkx2-1 (Nkx2-1) mRNA,
            complete cds.
ACCESSION   AF281080
VERSION     AF281080.1  GI:10442649
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1614)
  AUTHORS   Small,E.M., Vokes,S.A., Garriock,R.J., Li,D. and Krieg,P.A.
  TITLE     Developmental expression of the xenopus Nkx2-1 and Nkx2-4 genes
  JOURNAL   Mech. Dev. 96 (2), 259-262 (2000)
  MEDLINE   20417824
REFERENCE   2  (bases 1 to 1614)
  AUTHORS   Small,E.M., Vokes,S.A. and Krieg,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2000) Cell Biology and Anatomy, University of
            Arizona Health Sciences Center, 1501 N. Campbell Ave., Tucson, AZ
            85724, USA
FEATURES             Location/Qualifiers
     source          1. .1614
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /note="expressed in ventral forebrain, lungbud primordia
                     and thyroid anlage"
     gene            1. .1614
                     /gene="Nkx2-1"
     CDS             162. .1205
                     /gene="Nkx2-1"
                     /note="thyroid transcription factor; TTF-1; T/ebp; chicken
                     and mouse Nkx2-1 ortholog; contains homeodomain"
                     /codon_start=1
                     /product="homeobox transcription factor Nkx2-1"
                     /protein_id="AAG17405.1"
                     /db_xref="GI:10442650"
                     /translation="MSMSPKHTTPFSVSDILSPLEESYKKVAMEGAGLGAPLTAAAYR
                     QSQVSQASMQQHGMGHNGPVSAAYHMTAAGVPQLSHTTMGGYCNGNLGNLGNMSELPP
                     YQDTMRNSSATGWYGANPDPRFSTISRFMGPSGGMNMGGLGNMGSLGDVGKSMTPLQA
                     TPRRKRRVLFSQAQVYELERRFKQQKYLSAPEREHLASMIHLTPTQVKIWFQNHRYKM
                     KRQAKDKASQQQIQQDNSSCQQQQSPRRVAVPVLVKDGKPCQAGSNTPTAALQSHQQQ
                     TATAITITTNGLGPHQSHQTNSAGQSPDLVPHSNSPSSLQNQVTSLSHLNSSSSDYGS
                     AMSCSTLLYGRTW"
BASE COUNT      475 a    407 c    399 g    333 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 5;  Length 1614;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaagcg 25
           ||||||||||||||||
Db    717 TCTGCTGTTTGAAGCG 702


RESULT  37
AR120145
LOCUS       AR120145     5000 bp    DNA             PAT       16-MAY-2001
DEFINITION  Sequence 1 from patent US 6153741.
ACCESSION   AR120145
VERSION     AR120145.1  GI:14102844
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5000)
  AUTHORS   Richards,E.J. and Jeddeloh,J.A.
  TITLE     DNA methylation gene from plants
```

```
  JOURNAL   Patent: US 6153741-A 1 28-NOV-2000;
FEATURES             Location/Qualifiers
     source          1. .5000
                     /organism="unknown"
BASE COUNT     1386 a    909 c   1029 g   1676 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 6;  Length 5000;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctgtt 18
          ||||||||||||||||
Db   1132 TGCTGGATCTGCTGTT 1147


RESULT  38
AF143940
LOCUS       AF143940     5000 bp    DNA             PLN       11-MAY-1999
DEFINITION  Arabidopsis thaliana SWI2/SNF2-like protein (DDM1) gene, complete
            cds; and tRNA-Glu gene, complete sequence.
ACCESSION   AF143940
VERSION     AF143940.1  GI:4726078
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 5000)
  AUTHORS   Jeddeloh,J.A., Stokes,T.L. and Richards,E.J.
  TITLE     Maintenance of genomic methylation requires a SWI2/SNF2-like
            protein
  JOURNAL   Nat. Genet. 22 (1), 94-97 (1999)
  MEDLINE   99251588
REFERENCE   2  (bases 1 to 5000)
  AUTHORS   Jeddeloh,J.A. and Richards,E.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1999) Biology, Washington University, One
            Brookings Drive, St. Louis, MO 63130, USA
FEATURES             Location/Qualifiers
     source          1. .5000
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
     mRNA            join(<535. .566,772. .850,986. .1252,1354. .1440,1549. .1895,
                     1976. .2165,2251. .2426,2559. .2625,2703. .2892,2975. .3070,
                     3148. .3242,3317. .3436,3540. .3659,3745. .3843,3934. .4038,
                     4130. .>4354)
                     /gene="DDM1"
                     /product="SWI2/SNF2-like protein"
     gene            <535. .>4354
                     /gene="DDM1"
     CDS             join(535. .566,772. .850,986. .1252,1354. .1440,1549. .1895,
                     1976. .2165,2251. .2426,2559. .2625,2703. .2892,2975. .3070,
                     3148. .3242,3317. .3436,3540. .3659,3745. .3843,3934. .4038,
                     4130. .4354)
                     /gene="DDM1"
                     /codon_start=1
                     /product="SWI2/SNF2-like protein"
                     /protein_id="AAD28303.1"
                     /db_xref="GI:4726079"
                     /translation="MVSLRSRKVIPASEMVSDGKTEKDASGDSPTSVLNEEENCEEKS
                     VTVVEEEILLAKNGDSSLISEAMAQEEEQLLKLREDEEKANNAGSAVAPNLNETQFTK
                     LDELLTQTQLYSEFLLEKMEDITINGIESESQKAEPEKTGRGRKRKAASQYNNTKAKR
                     AVAAMISRSKEDGETINSDLTEEETVIKLQNELCPLLTGGQLKSYQLKGVKWLISLWQ
                     NGLNGILADQMGLGKTIQTIGFLSHLKGNGLDGPYLVIAPLSTLSNWFNEIARFTPSI
                     NAIIYHGDKNQRDELRRKHMPKTVGPKFPIVITSYEVAMNDAKRILRHYPWKYVVIDE
                     GHRLKNHKCKLLRELKHLKMDNKLLLTGTPLQNNLSELWSLLNFILPDIFTSHDEFES
                     WFDFSEKNKNEATKEEEEKRRAQVVSKLHGILRPFILRRMKCDVELSLPRKKEIIMYA
                     TMTDHQKKFQEHLVNNTLEAHLGENAIRGQGWKGKLNNLVIQLRKNCNHPDLLQGQID
                     GSYLYPPVEEIVGQCGKFRLLERLLVRLFANNHKVLIFSQWTKLLDIMDYYFSEKGFE
                     VCRIDGSVKLDERRRQIKDFSDEKSSCSIFLLSTRAGGLGINLTAADTCILYDSDWNP
                     QMDLQAMDRCHRIGQTKPVHVYRLSTAQSIETRVLKRAYSKLKLEHVVIGQGQFHQER
                     AKSSTPLEEEDILALLKEDETAEDKLIQTDISDADLDRLLDRSDLTITAPGETQAAEA
                     FPVKGPGWEVVLPSSGGMLSSLNS"
     variation       785
                     /gene="DDM1"
                     /note="site of ddm1-5 (som8) mutation; replace with 83 bp"
                     /replace=""
     variation       2384. .2385
                     /gene="DDM1"
                     /note="site of ddm1-6 (som4) mutation"
                     /replace=""
     misc_feature    3186
                     /gene="DDM1"
                     /note="alternate splice donor site used in ddm1-2"
     variation       3243
                     /gene="DDM1"
                     /note="site of ddm1-2 mutation"
                     /replace="a"
     variation       3337
                     /gene="DDM1"
                     /note="site of ddm1-7 (som5) mutation"
                     /replace="a"
     tRNA            complement(4755. .4826)
                     /product="tRNA-Glu"
                     /evidence=not_experimental
BASE COUNT     1386 a    909 c   1029 g   1676 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 8;  Length 5000;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctgtt 18
          ||||||||||||||||
Db   1132 TGCTGGATCTGCTGTT 1147


RESULT  39
YLI012084
LOCUS       YLI012084    6887 bp    DNA             PLN       23-SEP-1999
DEFINITION  Yarrowia lipolytica pex10 gene for peroxin 10.
ACCESSION   AJ012084
VERSION     AJ012084.1  GI:5921565
KEYWORDS    peroxin 10; pex10 gene.
SOURCE      Yarrowia lipolytica.
  ORGANISM  Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1  (bases 1 to 6887)
  AUTHORS   Le Dall,M.T., Gaillardin,C. and Nicaud,J.M.J.
  TITLE     PEX10 is essential for peroxisome biogenesis in Yarrowia lipolytica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6887)
  AUTHORS   Nicaud,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-1999) Nicaud J.M., Laboratoire de Genetique des
            Microorganismes, INRA-CNRS, BP 01, Tiverval-Grignon, F78850, FRANCE
FEATURES             Location/Qualifiers
     source          1. .6887
                     /organism="Yarrowia lipolytica"
                     /strain="W29"
                     /db_xref="taxon:4952"
     gene            complement(2293. .3357)
                     /gene="pex10"
     CDS             complement(2293. .3357)
                     /gene="pex10"
                     /function="essential for peroxisome biogenesis"
                     /codon_start=1
                     /evidence=experimental
                     /product="peroxin 10"
```

```
                     /protein_id="CAB56484.1"
                     /db_xref="GI:5921566"
                     /translation="MSDNTTIKKPIRPKPIRTERLPYAGAAEIIRANQKDHYFESVLE
                     QHLVTFLQKWKGVRFIHQYKEELETASKFAYLGLCTLVGSKTLGEEYTNLMYTIRDRT
                     ALPGVVRRFGYVLSNTLFPYLFVRYMGKLRAKLMREYPHLVEYDEDEPVPSPETWKER
                     VIKTFVNKFDKFTALEGFTAIHLAIFYVYGSYYQLSKRIWGMRYVFGHRLDKNEPRIG
                     YEMLGLLIFARFATSFVQTGREYLGALLEKSVEKEAGEKEDEKEAVVPKKKSSIPFIE
                     DTEGETEDKIDLEDPRQLKFIPEASRACTLCLSYISAPACTPCGHFFCWDCISEWVRE
                     KPECPLCRQGVREQNLLPIR"
BASE COUNT     1793 a    1841 c    1485 g    1768 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 8;  Length 6887;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 gatctgctgtttgaag 23
           ||||||||||||||||
Db    4051 GATCTGCTGTTTGAAG 4066


RESULT  40
AE004322/c
LOCUS       AE004322    11328 bp    DNA             BCT       31-JUL-2000
DEFINITION  Vibrio cholerae chromosome I, section 230 of 251 of the complete
            chromosome.
ACCESSION   AE004322 AE003852
VERSION     AE004322.1  GI:9657104
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 11328)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C. and White,O.
  TITLE     DNA sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
  JOURNAL   Nature 406 (6795), 477-483 (2000)
  MEDLINE   20406833
REFERENCE   2  (bases 1 to 11328)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11328
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /serotype="O1"
                     /db_xref="taxon:666"
                     /chromosome="I"
                     /note="biotype: El Tor"
     gene            complement(95. .286)
                     /gene="VC2521"
     CDS             complement(95. .286)
                     /gene="VC2521"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF95663.1"
                     /db_xref="GI:9657105"
                     /translation="MAANFDVLGIKREVEVPINYPYPLAILPVCHSKRQPLKLVPRLS
                     SACYRLCWMKTFIGFLDAK"
     gene            524. .1522
                     /gene="VC2522"
     CDS             524. .1522
                     /gene="VC2522"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF95664.1"
                     /db_xref="GI:9657106"
                     /translation="MAFIFIYYVGINMLIEAVVFLFIGLAILVWGADKLVFGSAALAR
                     NIGISPLVIGMTILAMGSSAPEMMVSATAALEGKTDTAVGNVLGSNIANIALILGITA
                     IVKPLSVGSAVLRRELPMMIGVTLIAGAILWDNHLGFYEGILLIALFAAFILVMLQVS
                     RKEKQNGDALLDEQESEIPVGVSNPKAAFWVVVGLILLPIGAGMLVDNAVVIAKYFGM
                     SDLVIGLTIIAVGTSLPELAASLAGALKGEDDMAVGNIIGSNVFNILAVMGLPGLLNP
                     SLLSPEAMSRDFWVMLGVSLLLVGVALGKKRQITRLEGALLLCSFIGYITYLLMNI"
     gene            1572. .2552
                     /gene="VC2523"
     CDS             1572. .2552
                     /gene="VC2523"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF95665.1"
                     /db_xref="GI:9657107"
                     /translation="MRMPHSFDYQTVAKQVLATEIHALQQLEQYINDDFARACAMILA
                     NQTGKVVVMGMGKSGHIGKKIAATLASTGTSAFFVHPGEASHGDLGMIERGDIVLAIS
                     NSGESSEILALLPVLKRLSIRVISMTGNPNSNMAKLADIHLQITVPREACPLELAPTS
                     STTATLVMGDALAVALMQARGFTAEDFALSHPGGALGRKLLLKLNDIMHSGDALPKVA
                     PQALIRDALLEISQKGLGMTAIVDEQDTLLGIFTDGDLRRILDKRIDIHSTVIADVMT
                     RQPTVAQPNLLAVEGLNLMQAKRINGLMLVENNKLVGALNMHDLLKAGVM"
     gene            2552. .3109
                     /gene="VC2524"
     CDS             2552. .3109
                     /gene="VC2524"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF95666.1"
                     /db_xref="GI:9657108"
                     /translation="MSSTVSTLYGEVEPSLLEIAKQIKLLICDVDGVFSDGLIYMGNQ
                     GEELKTFHTRDGYGVKALMNAGIEIAIITGRRSQIVENRMKALGISLIYQGQDDKVQA
                     YYDICQKLAIAPEQTGYIGDDLIDWPVMEKVALRVCVADGHPLLAQRANYVTHIKGGH
                     GAVREVCDLILQARNELDVHKGLSL"
     gene            3106. .3669
                     /gene="VC2525"
     CDS             3106. .3669
                     /gene="VC2525"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF95667.1"
                     /db_xref="GI:9657109"
                     /translation="MSLSRIVYVLLLFIASWSLYYLLGQEQDSKIQVAPNLELPMFSG
                     ENLENISYDEQGIRNYVITSIHLDHYAKSGNTLFKAPILKVYREGTLQEWEITARRGI
                     LSKDQVLTLYDDVLAKNLLPDSGFDTLTTSEMSIQLKSRDFWADKPVELRGPQFETHG
                     QAMKGNFADHSAELYNQVQGRYETLTP"
     gene            3605. .4144
                     /gene="VC2527"
     CDS             3605. .4144
                     /gene="VC2527"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF95668.1"
                     /db_xref="GI:9657110"
                     /translation="MPIIRLNFITKYKVDMKRSHLSLIACLLASTQAYALSTDREQPV
```

```
                YIDSDSQQLDMKSNRVTFLGDVKLKQGSININADKLIVIRNAENGKIEEIEGYGNVAT
                FSQLTDEGKTLYGEAKELYYKVRDDELVMINKAMLSQDDSEIRGSKIRYKISLQKLIA
                DSDGKGRVSTVLQPQAIQE"
gene            4147. .4872
                /gene="VC2528"
CDS             4147. .4872
                /gene="VC2528"
                /note="similar to PID:2317737; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="ABC transporter, ATP-binding protein"
                /protein_id="AAF95669.1"
                /db_xref="GI:9657111"
                /translation="MAILKAQHLAKSYKKRKVVSDVSLQVESGQIVGLLGPNGAGKTT
                SFYMIVGLVARDEGTITIDDNDISILPMHSRSRMGIGYLPQEASIFRKLSVEDNIMAV
                LQTREELTHEERQDKLEDLLEEFHIQHIRKSAGMALSGGERRRVEIARALAANPQFIL
                LDEPFAGVDPISVIDIKKIIEHLRDRGLGVLITDHNVRETLDVCEKAYIVSQGRLIAE
                GTPQDVLNNEQVKQVYLGEQFRL"
gene            4928. .6391
                /gene="VC2529"
CDS             4928. .6391
                /gene="VC2529"
                /note="similar to PID:2317738; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="RNA polymerase sigma-54 factor"
                /protein_id="AAF95670.1"
                /db_xref="GI:9657112"
                /translation="MKPSLQLKLGQQLAMTPQLQQAIRLLQLSTLDLQQEIQEALESN
                PLLEVEENQDEAPSLDSVRMTEESPREPEELYEPEPQDSSDLIEKSEISAELEMDTTW
                DEVYSANTGSTGLALDDDAPIYQGETTQTLQDYLHWQLDLTPFSDVDRTIAVALIDAI
                DDYGYLTVSLEEIQESLRSDDIELDEIEAVRKRIQQFDPFGVASLNLQDCLLLQLTTY
                PCDTPWLEEARLLLSQYIDDLGNRDYKTILKETKLKEEDLREILQLIQQLDPRPGSRI
                AQDHAEYVIPDVSVYKEQQGRWLVTINPDSVPKLKINQQYADLMRGNNAESNYIRTNLQ
                EAKWLIKSLESRNETLLKVAKCIVEHQHDFFEYGEEAMKPMVLNDVAMAVEMHESTIS
                RVTTQKYMHTPRGIFELKYFFSSHVSTDNGGECSSTAIRALIKKLVAAENPAKPLSDS
                KIATLLADQGIQVARRTIAKYRESLGIAPSSQRKRLL"
gene            6416. .6703
                /gene="VC2530"
CDS             6416. .6703
                /gene="VC2530"
                /note="similar to PID:2317739; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="sigma-54 modulation protein, putative"
                /protein_id="AAF95671.1"
                /db_xref="GI:9657113"
                /translation="MQINIQGHHIDLTDSMQDYVHSKFDKLERFFDHINHVQVILRVE
                KLRQIAEATLHVNQAEIHAHADDENMYAAIDSLVDKLVRQLNKHKEKLSSH"
gene            6706. .7152
                /gene="VC2531"
CDS             6706. .7152
                /gene="VC2531"
                /note="similar to PID:2317740; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="PTS system, nitrogen regulatory IIA component"
                /protein_id="AAF95672.1"
                /db_xref="GI:9657114"
                /translation="MQLSEVLSLDCTKSAVQCSSKKRALEIISEIAALHTGQNATELF
                ECMLSREKMGSTGIGNGIAIPHARMSDSENAVAVLLQCEEPIEFDSIDNRPVDLLFAL
                LVPEDQCKEHLKTLACMAERLNDKQILKQLRNAASDQELYNIMVNQ"
gene            7165. .8025
                /gene="VC2532"
CDS             7165. .8025
                /gene="VC2532"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF95673.1"
                /db_xref="GI:9657115"
                /translation="MRLIVVSGQSGAGKSVALRVLEDLGYYCVDNLPVSLLTAFIQSV
                QGSQQNVAVSIDIRNLPKEPSLVQDVLDQLKQNNDVSMLFLDASKETLLKRYSETRRI

  Query Match             64.0%;  Score 16;  DB 1;  Length 11328;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tggatctgctgtttga 21
           ||||||||||||||||
Db    7838 TGGATCTGCTGTTTGA 7823


RESULT  41
AE004026
LOCUS       AE004026    13856 bp    DNA             BCT       15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 172 of 229 of the complete genome.
ACCESSION   AE004026 AE003849
VERSION     AE004026.1  GI:9107228
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 13856)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
   PUBMED   10910347
REFERENCE   2  (bases 1 to 13856)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  T:        Direct Submission
  JO..NAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
```

```
          Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
          13083-970, Brazil
FEATURES             Location/Qualifiers
     source          1. .13856
                     /organism="Xylella fastidiosa 9a5c"
                     /db_xref="taxon:160492"
                     /clone="9a5c"
     gene            complement(195. .803)
                     /gene="XF2101"
     CDS             complement(195. .803)
                     /gene="XF2101"
                     /note="similar to GI|1652989 (percent identity: 30 %/query
                     alignment coverage:  87.6 %/subject alignment coverage:
                     67.6 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF84900.1"
                     /db_xref="GI:9107229"
                     /translation="MSRCITGYRWWQWICVVVVLVFFWLVAVAGWIVWIGQRDQAASS
                     DVIIVLGAAAYDAKPSPVFEERIRHGLYLYQHGYAPTLIFTGGFGGSSARFAESQVAR
                     RYALRYDVPPRAILIETVSRTTRQNLIQARQLMRVYGLRRAIVVSDPLHMARALRLCR
                     ELDIDALASATPSTRFRSLHTRWRFLLQEVYFFHRDLVIQAN"
     gene            complement(800. .1042)
                     /gene="XF2102"
     CDS             complement(800. .1042)
                     /gene="XF2102"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84901.1"
                     /db_xref="GI:9107230"
                     /translation="MARVIMFYSSFALSVKSCIGVLTGMLLLHAYVFGCGLGGEFRAL
                     CCDVAWLCHVDISARLSGCGCSAGFFICGVSVLVVS"
     gene            1397. .1732
                     /gene="XF2103"
     CDS             1397. .1732
                     /gene="XF2103"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84902.1"
                     /db_xref="GI:9107231"
                     /translation="MKARLFLPILIMSAISGCATYTYSNGYYHGTPTVEYHYPPNYYY
                     PHGYGMHDLFNYPIPPIYPRDFYRTHIHLPYWQRPHHDPSRPLPPKQPIKQEQPNHIE
                     MGLRDKQPI"
     gene            1914. .2024
                     /gene="XF2104"
     CDS             1914. .2024
                     /gene="XF2104"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84903.1"
                     /db_xref="GI:9107232"
                     /translation="MNIADQDLMLLCSELELMKICPHCVDTLKHQRLAAR"
     gene            complement(2067. .2225)
                     /gene="XF2105"
     CDS             complement(2067. .2225)
                     /gene="XF2105"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84904.1"
                     /db_xref="GI:9107233"
                     /translation="MAAWSREEAWMEFYVFCVIFRNALFRVGTFYAVDVVAPAAALRL
                     QAQRSGLE"
     gene            2271. .4160
                     /gene="XF2106"
     CDS             2271. .4160
                     /gene="XF2106"
                     /note="similar to SP|P25756 (percent identity: 65 %/query
                     alignment coverage:  98.6 %/subject alignment coverage:
                     98.4 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose inhibited division protein A"
                     /protein_id="AAF84905.1"
                     /db_xref="GI:9107234"
                     /translation="MTDSFYGYDVIVIGAGHAGSEAALAAARTGAHTLLLTHNIETIG
                     AMSCNPAIGGIGKGHLVKEIDALGGAMAHAADAAGIQWRTLNASKGPAVRATRCQADR
                     ALYRAAIRQMIEGQARLNIFQAEVDDLLFEGDTVCGAITHTGLHFKAPAVVLTAGTFL
                     AGKIHIGPTQYAAGRMGDPPATMLAARLRERLFTVARLKTGTPPRIDGRTLNYTAMQE
                     QPGDAPRPTMSFIGNSASHPPQVSCWITQTTERTHEIIRAALHRSPLYSGQIEGAGPR
                     YCPSIEDKVVRFAEKNSHQIFVEPEGLNVIDIYPNGISTSLPFDVQLELVRSIRGFEQ
                     AHITRPGYAIEYDFFDPRGLKASLETKAIAGLFFAGQINGTTGYEEAAAQGLLAGLNA
                     ARHVRGLSSWTPRRDQAYLGVLVDDLITHGTNEPYRMFTSRAEYRLQLREDNADARLT
                     AIGRDLGLIDDARWAHFNAKQEAVARECERLSAFWATPGNALGREVKETLGVTLSRET
                     NIIDLMKRPELDYAALMRVPSLGPGVDDAQVAEQVEISVKYAGYLNRQSEEITRQQRH
                     EATAIPLEFDYAAVRGLSTEVLQKLQHIQPQTVGQAQRIPGMTPAAISLLLVHLERMR
                     RNRVA"
     gene            complement(4348. .4560)
                     /gene="XF2107"
     CDS             complement(4348. .4560)
                     /gene="XF2107"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84906.1"
                     /db_xref="GI:9107235"
                     /translation="MFILLMCVHTHPLIDELVEAWRAGPVIRSMYNAIKQDGSSGITE
                     RLPVRWFSWGRASKIDATAAAILASV"
     tRNA            4928. .5001
                     /gene="tRNA-Cys"
                     /note="found by tRNAscan"
                     /product="tRNA-Cys"
     gene            4928. .5001
                     /gene="tRNA-Cys"
     gene            5207. .5392
                     /gene="XF2108"
     CDS             5207. .5392
                     /gene="XF2108"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84907.1"
                     /db_xref="GI:9107236"
                     /translation="MQADGTTIIGLDISTKRWATAFLSGQNPLSIEDKSDFVLGAALR
                     NSAKNDTTQRLAPIKTA"
     gene            complement(5566. .5892)
                     /gene="XF2109"
     CDS             complement(5566. .5892)
                     /gene="XF2109"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84908.1"
                     /db_xref="GI:9107237"

Query Match          64.0%;  Score 16;  DB 1;  Length 13856;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 gatctgctgtttgaag 23
           ||||||||||||||||
Db    2652 GATCTGCTGTTTGAAG 2667


RESULT  42
AE004580
LOCUS       AE004580    15247 bp    DNA             BCT       30-AUG-2000
DEFINITION  Pseudomonas aeruginosa PA01, section 141 of 529 of the complete
            genome.
ACCESSION   AE004580 AE004091
VERSION     AE004580.1  GI:9947468
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa.
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Pseudomonas.
REFERENCE   1  (bases 1 to 15247)
  AUTHORS   Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
            Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
            Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
            Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
REFERENCE   2  (bases 1 to 15247)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
FEATURES             Location/Qualifiers
     source          1. .15247
                     /organism="Pseudomonas aeruginosa"
                     /strain="PAO1"
                     /db_xref="taxon:287"
     gene            complement(122. .640)
                     /gene="hc"
                     /note="PA1512"
     CDS             complement(122. .640)
                     /gene="hc"
                     /codon_start=1
                     /transl_table=11
                     /product="secreted protein Hcp"
                     /protein_id="AAG04901.1"
                     /db_xref="GI:9947469"
                     /translation="MATPAYMSITGTKQGLITAGAFTEDSVGNTYQEGHEDQVMVQGF
                     NHEVIIPRDPQSGQPTGQRVHKPVVITKVFDKASPLLLAALTSGERLTKVEIQWYRTS
                     AAGTQEHYYTTVLEDAIIVDIKDYMHNCQDPGNAHFTHLEDVHFTYRKITWTHEVSGT
                     SGSDDWRSPVAG"
     gene            complement(1126. .2403)
                     /gene="PA1513"
     CDS             complement(1126. .2403)
                     /gene="PA1513"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04902.1"
                     /db_xref="GI:9947470"
                     /translation="MEAHLIEWLNLLVRWIHMIVGIAWIGASFYFVWLENNLNRSNPR
                     EGLSGDLWAIHGGGIYHLEKYKLAPPKMPENLHWFKWEAYSTWMSGVVLLTIVFYLNP
                     TLYLIAPGSDLAPAAAIAIGIGSLVCGWFVYTILCDSALGKTPALLGLVLFVLLVAAA
                     YGLTQVFSGRGAYLHVGAIIGTIMVGNVFRVIMPAQRALVKAIEENREPDPVLPAKGL
                     LRSRHNNYFTLPVLFIMISNHFPSTYGSQYNWLILAAIAVLAVLVRHYFNTRHESNRF
                     AWALPAGALGMIALAFVTTPNYQQASAPLASTPAAAPAQPATGGDAPAAPAATQASNA
                     GAASGADFDKVHHVIQERCTVCHSAKPTSQLFSTAPGGIMLDTPQQIQQLAPKIQAQA
                     VASQIMPLGNITQMTQEERNLIGSWIAKGAQIN"
     gene            complement(2452. .2961)
                     /gene="PA1514"
     CDS             complement(2452. .2961)
                     /gene="PA1514"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG04903.1"
                     /db_xref="GI:9947471"
                     /translation="MRTLKIEPLTKEAFAPFGDVIETAGSDYFMINNGSTRRYHKLAT
                     VETAQPEDNAIISIFSAEKLEMPLRIRMLERHPLGSQAFIPLLGNPFLVVVAPLGDVP
                     VPGLVRAFLTNGRQGVNYHRGVWHHPVLTIEKRDDFLVVDRSGSGNNCDEHFFTEDEQ
                     LLLDPQSNQ"
     gene            complement(3016. .4014)
                     /gene="alc"
                     /note="PA1515"
     CDS             complement(3016. .4014)
                     /gene="alc"
                     /codon_start=1
                     /transl_table=11
                     /product="allantoicase"
                     /protein_id="AAG04904.1"
                     /db_xref="GI:9947472"
                     /translation="MNAEHSPFRHYLDLADARLGSQVVAVSDEWFAPASRMLQAGEPV
                     WKEGVFDDSGKWMDGWETRRKRFEGHDQAVIRLGVPGVLKGVDIDTRFFTGNHPPAAS
                     LDGCFCAEGDPDDSTSWSEVLAAVGLQGDSHHYHPIDDERPWTHLRLNIYPDGGIARL
                     RLYGVPYRDWSNQPPGTALDLAAAVNGGRALACSDQHFGRMGNLLNPGRAINMGDGWE
                     TGRRRTPGHDWVIVALGHPGSIEAAVVDTLHFKGNYPESCSIQAAFVEGGNEARIEAQ
                     SLFWRELLPAQKLEMHQEHRFERHLNALGPITHVRLNIFPDGGVSRLRLFGRPQLP"
     gene            complement(4032. .4547)
                     /gene="PA1516"
     CDS             complement(4032. .4547)
                     /gene="PA1516"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAG04905.1"
                     /db_xref="GI:9947473"
                     /translation="MSRFQTLTPASLSREAFIEAFADIYEHSPWVAEKAYDLGIDDSL
                     NDIEGLHQRMADILLSASREAQLALINAHPDLAGKAAIRGELTASSTSEQAGAGIHEC
                     SAEEFARFTELNDAYKARFGFPFIKAVKGSNRHQILAAFEERIQHSADEEFATALAEI
                     NKIALFRLQQL"
     gene            complement(4544. .5470)
                     /gene="PA1517"
     CDS             complement(4544. .5470)
                     /gene="PA1517"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG04906.1"
                     /db_xref="GI:9947474"
                     /translation="MSADYPRDLIGYGNNPPHPHWPGDARIALSFVLNYEEGGERCVL
                     HGDKESEAFLSEMVAAQPLQGVRHMSMESLYEYGSRAGVWRLLKLFKRRNVPLTVFAV
                     AMAAQRNPEVIRAMVADGHEICSHGYRWIDYQYMDEAQEREHMLEAIRILTELTGQRP
                     VGWYTGRTGPNTRRLVMEEGGFLYDSDTYDDDLPYWDPASTAEKPHLVIPYTLDTNDM
                     RFTQVQGFNNGEQFFQYLKDAFDVLYEEGATAPKMLSIGLHCRLIGRPARMAALERFI
                     QYAQSHDKVWFARREDIARHWHREHPFQETEA"
     gene            5843. .6223
                     /gene="PA1518"
     CDS             5843. .6223
                     /gene="PA1518"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAG04907.1"
                     /db_xref="GI:9947475"
                     /translation="MSASGRPQPMGRLTTHVLDSAHGCPGHGIKIELYRVEGQQLELI
                     ATTLTNHDGRCDQPVLQGEDFRPGVYQLVFNAGDYYRARGVQLPEPAFLDQVVLRFGI
```

```
                ASADDHYHVPLLISPYSYSTYRGS"
     gene       complement(6464. .7813)
                /gene="PA1519"
     CDS        complement(6464. .7813)
                /gene="PA1519"
                /codon_start=1
                /transl_table=11
                /product="probable transporter"
                /protein_id="AAG04908.1"
                /db_xref="GI:9947476"
                /translation="MESTKQEEQGIYAATPPATGLLERLFKLRQHGTTVRTELAAGLT
                TFITMAYIIFVNPNIMADAGIDHGAAFVATCLAAALGCFLMGLYANWPVGLAPGMGLN
                AFFTYTVVGTMGYSWQIALGAVFISGVMFMLLTFSRVREWLLNSIPRSLRFAMGAGVG
                LFLGLIGLKTAGIVVASPATLIKLGHLTSPGPLLAALCFLMIAVLEYRRVFGGILISI
                LSVTLVGWALGLVQYGGVFSAPPSLAPTFLAMDIAGAFNVTMISVILAFLFVHMFDTA
                GTLMGVAQRAHLVKEDGRIENLSKAMKADSASSVFGGMLGVPPVTSYVESAAGVAAGG
                RTGLTAVVVGVLFVAAMFFAPLAGMIPAFATAGALIYVAMLMMGGMAHIDWDEHTETI
                PAIVTVIMMPLTFSVADGIALGFITYVAMKVLTGKHREVSASLYALFVIFVAKFIFL"
     gene       8188. .8967
                /gene="PA1520"
     CDS        8188. .8967
                /gene="PA1520"
                /codon_start=1
                /transl_table=11
                /product="probable transcriptional regulator"
                /protein_id="AAG04909.1"
                /db_xref="GI:9947477"
                /translation="MTDQLQHIRKPPRTGKAARSGTQDEIVYAHIFDAILEQRLAPGT
                KLSEEALGEIFGVSRTIIRRALSRLAHEQVVLLRPNRGAVVASPSIDEARQILFARRT
                VERAITELATDNATAERLAELREMVVQEQSSFARGDRGAGIRLSGEFHLKLAEMARNA
                PLVSFQRSLVSQTSLIIAQYESGGRSHCSFDEHNEILDAIEKGDKERAVTLMMHHMEH
                IDSKLNLESDNASGDLHAVFSHLLGGKKKPRRAKADSDSAA"
     gene       complement(9021. .10325)
                /gene="PA1521"
     CDS        complement(9021. .10325)
                /gene="PA1521"
                /codon_start=1
                /transl_table=11
                /product="probable guanine deaminase"
                /protein_id="AAG04910.1"
                /db_xref="GI:9947478"
                /translation="MTRNAKAYRAAILHSIADPAEVGVERSYEYFEDGLLLVEDGKVA
                RLGDAETLLGEIGEVEVFEYRDALITPGFIDTHIHFPQTGMIASYGEQLLDWLNTYTF
                PTERQFGDQAHADQVAEIFLQELLRNGTTTALVFGSVHRQSVESLFEAARRLDLRLIA
                GKVMMDRNAPDYLTDTAESSYRDSKALIERWHGQGRLLYAVTPRFAPTSTAEQLDMAA
                RLLREHPGVYLHTHLSENLKEIEWVKELFPERSGYLDVYDHHGLLGPRSVFAHGVHLC
                DGECQRLAETGSAVAFCPTSNLFLGSGLFDLPKLERYKVKVGLGTDVGAGTSFSQLQS
                LNEAYKVMQLQGARLDPFKSLYLATLGGARALELDDRIGSFATSNEADFVVLDYHATP
                LLSYRLSQAGSLAERLFALTILGDDRTVKETFAAGRSVHRRD"
     gene       complement(10395. .11234)
                /gene="PA1522"
     CDS        complement(10395. .11234)
                /gene="PA1522"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAG04911.1"
                /db_xref="GI:9947479"
                /translation="MNWISALAELQRSGEPCVLVTIIEERGSTPRNAGSKMVVSAERL
                YDTIGGGHLEYRAQAIAREMLAARTQDTRLERFSLGASLGQCCGGATVLLFEPMGQPQ
                AHIAVFGAGHVGRALVPLLASLPCRVRWIDSREHEFPALLPDGVEKVVNDEVLDEVER
                MPPGSHFIVMTHNHPLDLELTAAILARNDFAYFGLIGSKTKRVKFEHRLRERGVDAER
                LQRMRCPMGLEQVKGKLPAEIAISIAGEVIATYNAAFGLERKQGPSSVTRLTPESRRA
                HES"

  Query Match             64.0%;  Score 16;  DB 1;  Length 15247;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db    1254 GTTGCTGGATCTGCTG 1269
```

```
RESULT  43
HS314G4
LOCUS       HS314G4     31557 bp    DNA             PRI       23-NOV-1999
DEFINITION  Human DNA sequence from cosmid 314G4 from a contig from the tip of
            the short arm of chromosome 16, spanning 2Mb of 16p13.3. Contains
            ESTs and CpG islands.
ACCESSION   Z69667
VERSION     Z69667.1  GI:1200312
KEYWORDS    16p13.3; CpG island.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 31557)
  AUTHORS   Thomas,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1996) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail enquires regarding this sequence:
            humquery@sanger.ac.uk
COMMENT     IMPORTANT:
            This sequence is not the entire insert of clone 314G4. It may be
            shorter because we only sequence overlapping sections once, or
            longer because we arrange for a small overlap between neighbouring
            submissions.
            The true right end of clone 314G4 is at 31557.
            314G4 is from a 280kb clone contig extending from the telomere of
            16p. Higgs D.R., Flint J. unpublished. MRC Molecular Haematology
            Unit, Institute of Molecular Medicine, Oxford.
            314G4 came from the Los Alamos, flow sorted human Chromosome 16
            libraries constructed by Norman Doggett (unpublished).
FEATURES             Location/Qualifiers
     source          1. .31557
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16p13.3"
                     /clone="LA16-314G4"
                     /clone_lib="LA16"
     repeat_region   428. .716
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    914. .1691
                     /note="expressed region - matched by multiple ESTs"
     misc_feature    1692. .2316
                     /partial
                     /note="match: 5' EST R51457 clone 39198"
     misc_feature    2290. .2810
                     /partial
                     /note="match: 5' EST T87094 clone 115315"
     misc_feature    complement(3171. .3720)
                     /note="expressed region - matched by multiple ESTs"
     repeat_region   5136. .5489
                     /note="3 copies of 118 mer 88 % conserved"
     misc_feature    5908. .6086
                     /note="match: EST Z39463 clone c-1ag06"
     repeat_region   7536. .7567
                     /partial
                     /note="Alu repeat: matches 1. .32 of consensus"
     repeat_region   7568. .7856
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   9446. .9734
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    12398. .13777
                     /note="putative CpG island"
     repeat_region   13241. .13275
                     /note="7 copies of 5 mer 86 % conserved"
     repeat_region   14561. .14677
                     /partial
                     /note="Alu repeat: matches 254. .136 of consensus"
     repeat_region   14680. .14971
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   14974. .15106
                     /partial
```

```
                     /note="Alu repeat: matches 145. .1 of consensus"
     repeat_region   15127. .15231
                     /partial
                     /note="Alu repeat: matches 116. .1 of consensus"
     repeat_region   15248. .15535
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   15607. .15917
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   16005. .16114
                     /partial
                     /note="Alu repeat: matches 108. .1 of consensus"
     repeat_region   16123. .16283
                     /partial
                     /note="Alu repeat: matches 308. .150 of consensus"
     misc_feature    17661. .18594
                     /note="putative CpG island"
     repeat_region   18912. .19019
                     /note="3 copies of 36 mer 89 % conserved"
     repeat_region   19514. .19621
                     /note="MER22 element fragment"
     misc_feature    19922. .20499
                     /note="match: 5' EST H21176 clone 173392"
     repeat_region   20504. .20539
                     /note="4 copies of 9 mer 97 % conserved"
     repeat_region   21472. .21766
                     /note="Alu repeat: matches 308. .1 of consensus"
     misc_feature    25040. .25378
                     /note="expressed region - matched by multiple ESTs"
     misc_feature    complement(25620. .25910)
                     /note="expressed region - matched by multiple ESTs"
     repeat_region   26707. .26817
                     /partial
                     /note="Alu repeat: matches 298. .184 of consensus"
     repeat_region   26818. .26963
                     /partial
                     /note="Alu repeat: matches 160. .1 of consensus"
     misc_feature    complement(27039. .27209)
                     /note="expressed region - matched by multiple ESTs"
     repeat_region   30474. .30611
                     /note="3 copies of 46 mer 94 % conserved"
     misc_feature    31087. .31220
                     /note="expressed region - matched by multiple ESTs"
     repeat_region   31496. .31557
                     /partial
                     /note="Alu repeat: matches 1. .62 of consensus"
BASE COUNT     5713 a   9969 c   9918 g   5957 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 9;  Length 31557;
  Best Local Similarity   100.0%;  Pred. No. 94;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db  10104 GTTGCTGGATCTGCTG 10119


RESULT  44
SC8520X
LOCUS       SC8520X     41200 bp    DNA             PLN       11-AUG-1997
DEFINITION  S.cerevisiae chromosome XIII cosmid 8520.
ACCESSION   Z49705 Z49700 Z71257
VERSION     Z49705.1  GI:825556
KEYWORDS    ALD2; aldehyde dehydrogenase; ATPase; CBF3B; centromere DNA-binding
            protein; coiled-coil; DNA mismatch repair; DnaJ; Mitochondrial
            carrier protein; MLH1; SMP2; transfer RNA-Met; transfer RNA-Val.
SOURCE      baker's yeast.
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 41200)
  AUTHORS   Hunt,S. and Bowman,S.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 41200)
  AUTHORS   Barrell,B. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-1995) Saccharomyces cerevisiae chromosome XIII
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT     Notes:
            All CDS over 100 codons have been analysed. CDS that are completely
            overlapped and those that are overlapped by more than 50%
            of their length by a larger CDS have been omitted from this
            analysis.
            Details of the omitted CDS are available on request. The more
            significant matches with motifs in the PROSITE database are
            also included but some of these may be fortuitous. The length in
            codons and the calculated codon adaptation index (CAI)
            is given for each CDS.
            Cosmid 8520 is overlapped at the start by cosmid 9375, EMBL id
            SC9375,
            accession no. Z47071 and at the end by by cosmid 8010.
FEATURES             Location/Qualifiers
     source          1. .41200
                     /organism="Saccharomyces cerevisiae"
                     /strain="AB972"
                     /db_xref="taxon:4932"
                     /chromosome="XIII"
                     /clone="cosmid 8520"
                     /map="13R"
     misc_feature    1. .365
                     /note="overlap with cosmid 9375"
     CDS             <1. .514
                     /note="YM8520.01, len: > 170, CAI: 0.25, overlaps an
                     extends YM9375.22, SW:YMAL_YEAST P28625, previously noted
                     hypothetical 41.6 KD protein in IMP1 3' region (RF1095),
                     hypothetical transmembrane protein"
                     /codon_start=2
                     /product="unknown"
                     /protein_id="CAA89788.1"
                     /db_xref="GI:825557"
                     /db_xref="SWISS-PROT:P28625"
                     /translation="VVGICSKNSIERNKKLGYDYLVPYDEGSIVENVKKLKQSVLEND
                     KFDMIFDSVGNHDFFPVIDQFLKPKAKNSFYVTIAGNNKADYKNISWRDFVSLSSILK
                     AINPFKKYNWRFGHPYPPNNFIEVGNEMIKKGTYKPPIDSVYEFDQYKEAIDRLMSNR
                     AKGKVVVKMK"
     CDS             756. .2183
                     /note="YM8520.02, unknown, len: 475, CAI: 0.12"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA89789.1"
                     /db_xref="GI:825558"
                     /db_xref="SWISS-PROT:Q03790"
                     /translation="MADLQKQENSSRFTNVSVIAPESQGQHEQQKQQEQLEQQKQPTG
                     LLKGLNGFPSAPQPLFMEDPPSTVSGELNDNPAWFNNPRKRAIPNSIIKRSNGQSLSP
                     VRSDSADVPAFSNSNGFNNVTFGSKKDPRILKNVSPNDNNSANNNAHSSDLGTVVFDS
                     NEAPPKTSLADWQKEDGIFSSKTDNIEDPNLSSNITFDGKPTATPSPFRPLEKTSRIL
                     NFFDKNTKTTPNTASSEASAGSKEGASTNWDDHAIIIFGYPETIANSIILHFANFGEI
                     LEDFRVIKDFKKLNSKNMSKSPSLTAQKYPIYTGDGWVKLTYKSELSKSRALQENGII
                     MNGTLIGCVSYSPAALKQLASLKKSEEIINNKTSSQTSLSSKDLSNYRKTEGIFEKAK
                     AKAVTSKVRNAEFKVSKNSTSFKNPRRLEIKDGRSLFLRNRGKIHSGVLSSIESDLKK
                     REQASKSKKSWLNRLNNWLFGWNDL"
     CDS             complement(2320. .4503)
                     /note="YM8520.03c, unknown, len: 727, CAI: 0.12, some
                     similarity to WP:T04A8.16 CE01070 calpain thiol protease,
                     (27.1% identity in 155 aa overlap)"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="CAA89790.1"
                     /db_xref="GI:825559"
                     /db_xref="SWISS-PROT:Q03792"
                     /translation="MNDWHEFNAAIKSIYCNAEGDSSSIINRLVGLAMKSEDSTFIEA
                     VLVLKENVSKVDKQLRFLWLTSTINSRFYPPIPISEASPVSWNKTEYCAPGTEELQRR
                     YPGRAKLQNEEDYSGGIEQCRDVPDCSLVASLINLRSKNLNLPLIKQISSTKYHVNLS
                     FNGSNKRLVTVDISQIPTSVDGKQLSLKSNDISDKIGELALLLVSKGTYSTDGSNISI
```

```
                DTYRLSGFLPEITQVNSYPFEKLWKFHKSNLCLMGAGTGNRSNDMIKPLVENHDYSII
                DITYDSRLVKLRDPRNSALNVEISYEQYLKNFKQLYLNWNQEKLFKRSQVLHFRYDTS
                RYNKFSIVADKPLFHLVNNSKVTETVWLLLESHLQDEGSQENRSVSFLNEAPECIICP
                IEPPVECGGNHIGLQLVKLRLDAETERLLYCYSTTNNNFSIHSFSVVKEICFQRLKDT
                KSLFAKVLFSFPYEIEGKASFDTCNFFQNPTFELEVHSEQDYQVLMDAACISTSSHDL
                INIQVYYFNDYELIKPIMFDNHYQPGQGLKQDVPILTNVKYMIVCSTYGPPASTEFEL
                LASIRLSSSWRLISGITLRSVNLIYGTYPYHCRNRFHWKETSDKLKIQMTLPTKKYST
                NKLFIRVVPVESSARLRMRCNIFEPESALCVYECQEYRTCPSGGIVIPDLEVSRTNIV
                VLMIERSVPISSCLPTEGQLDELELFVGSSQKIRIEKYSDDVIPK"
CDS             4872. .6515
                /note="YM8520.04, unknown, len: 547, CAI: 0.10, potential
                integral membrane protein"
                /codon_start=1
                /product="unknown"
                /protein_id="CAA89791.1"
                /db_xref="GI:825560"
                /db_xref="SWISS-PROT:Q03795"
                /translation="MVKKHQNSKMGNTNHFGHLKSFVGGNVVALGAGTPYLFSFYAPQ
                LLSKCHIPVSASSKLSFSLTIGSSLMGILAGIVVDRSPKLSCLIGSMCVFIAYLILNL
                CYKHEWSSTFLISLSLVLIGYGSVSGFYASVKCANTNFPQHRGTAGAFPVSLYGLSGM
                VFSYLCSKLFGENIEHVFIFLMVACGCMILVGYFSLDIFSNAEGDDASIKEWELQKSR
                ETDDNIVPLYENSNDYIGSPVRSSSPATYETYALSDNFQETSEFFALEDRQLSNRPLL
                SPSSPHTKYDFEDENTSKNTVGENSAQKSMRLHVFQSLKSSTFIGYYIVLGILQGVGL
                MYIYSVGFMVQAQVSTPPLNQLPINAEKIQSLQVTLLSLLSFCGRLSSGPISDFLVKK
                FKAQRLWNIVIASLLVFLASNKISHDFSSIEDPSLRASKSFKNISVCSAIFGYSFGVL
                FGTFPSIVADRFGTNGYSTLWGVLTTGGVFSVSVFTDILGRDFKANTGDDDGNCKKGV
                LCYSYTFMVTKYCAAFNLLFVLGIIGYTYYRRRATANSL"
CDS             complement(6621. .7337)
                /note="YM8520.05c, unknown, len: 238, CAI: 0.15"
                /codon_start=1
                /product="unknown"
                /protein_id="CAA89792.1"
                /db_xref="GI:825561"
                /db_xref="SWISS-PROT:Q03796"
                /translation="MSHKLTILPFLIKFTPKFPQSIDHDEHGLNVYAFDLDHTIIKPK
                SPNISFSRSASDWQFINFNSKKSTLDYLCNIIDNDPTAVIVIFSNQGGVITVPRTSKS
                CTKYTNKILLFLKAIKNDERGETLSHRLWLYAAPKRPKTFAANHSKITFASLGESYNN
                DPNIFEKVRKPMTGMVEFFKRDLESAYRVSEQISPIKLNWIYYCGDAAGRKKDFSDSD
                IKFAENLHVEFKYPEEIFHG"
CDS             complement(7598. .8365)
                /note="YM8520.06c, unknown, len: 255, CAI: 0.17"
                /codon_start=1
                /product="unknown"
                /protein_id="CAA89793.1"
                /db_xref="GI:825562"
                /db_xref="SWISS-PROT:Q03798"
                /translation="MLRPLRKSVLASCRHCFKVCGGLPQKQLPLFSPLLLRARYSSTD
                SSTKRSNKSDKIDAPGFKKIFLVAIIGTVIFVKTVQSLDKNKPKTTLSEEEFENVVKG
                LKRRVAIFPQGEVDIKFSLSPSIEETRKVLQKSQGDDINELQFVDPVKVIDYYRTLRD
                DRYEALLNEYYKKYGCDTYAYNLPTGMLVMLLGRYFKENFKTGDKLVVVNFPHSIADA
                TRFENEVSIVSKIFVPRKLSGSDVCKYYETVGKADII"
CDS             8569. .9036
                /note="YM8520.07, unknown, len; 155, CAI: 0.13"
                /codon_start=1
                /product="unknown"
                /protein_id="CAA89794.1"
                /db_xref="GI:825563"
                /db_xref="SWISS-PROT:Q03799"
                /translation="MSLVKLANTCAHLQNCSKVRVALTSIPYTKLQLQFAYNLYQQGF
                LSSLQKGSTMGPDKDFVEVTPDNISTRRLWVGLKYRDNKPVLSSCKLISKPNSRIHLP
                MEDMKKLCSGVTIRNIKPLQPGELILVRAHNNIMDINEAISKKLDGEVLCRVK"
tRNA            complement(9204. .9276)
                /gene="tRNA-Met"
                /note="tRNA-Met gene, anticodon cat, len: 73"
gene            complement(9204. .9276)
                /gene="tRNA-Met"
LTR             complement(9432. .9659)
                /note="remnant of solo delta element"
CDS             complement(10797. .11249)
                /note="YM8520.08c, unknown, len: 150, CAI: 0.13, possible
                coiled-coil protein"
                /codon_start=1
                /product="unknown"
                /protein_id="CAA89795.1"
                /db_xref="GI:825564"
                /db_xref="SWISS-PROT:Q03818"
                /translation="MGNFIITERKKAKEERSNPQTDSMDDLLIRRLTDRNDKEAHLNE
                LFQDNSGAIGGNIVSHDDALLNTLAILQKELKSKEQEIRRLKEVIALKNKNTERLNDE
                LISGTIENNVLQQKLSDLKKEHSQLVARWLKKTEKETEAMNSEIDGTK"
CDS             11387. .13837
                /note="YM8520.09, unknown, len: 816, CAI: 0.13"
                /codon_start=1
                /product="unknown"
                /protein_id="CAA89796.1"
                /db_xref="GI:825565"
                /db_xref="SWISS-PROT:Q03823"
                /translation="MDVVQQKQDKQLQHQTQEQQQIREDQQEVPPQRPRQQNRWKPWW
                NSTADDEPNTGRIAEYPNGQGRSSPTTDFQDSVNSNNDNKGIWSKIASFATSRYRSAP
                IVVDDNTRYSQLNTEQINFLENEAKDIISKKSKSWCWYEAIPHISNSSNIIDSIDTPG
                IISVSGTGSAKCPLPLNKYPGEGGNPGYNVFINDSLILPSDNPLNFLHVQPLRTKVLN
                TIKNYYNFPNEQHLYLRQKKTALLKDKRIIIISVVGDLPEKYEQRSLESQRSAYYLSR
                KLSQNLAQEQPQRVLTLSFQCPLHNQDLIPTYKECVELLNHWAHLFKEVDSIFFVGVY
                HSVPLTLLLAKYIVQNNEVLEFDENTTVSVLSFQSCLQGYRFWDHSTDFTSNSYNNLG
                SNSSTNENDSNDHDSNNDFTTKSQQIKEKQLFQGIDKKQQDTLSKIKNYRRIDSSESK
                LVQDALDWLLFNWDTFRLTFFGKLYDNFMTISEKLAIDYNHPKILRNLWCNGKYMGID
                LKNANNLNLDTDDEATSNINDVHVRTPNFESRLKIPTNRLFEITLWDILMITENLGYK
                QFIPIINLLSPFFISRSFNDYTLPPNIRKQYQNSNKIWLQEMDSKWKMNGHQLNYDQR
                EGESLGSSSESLLPENISTVKDFLQFVQYQNEKSSDFVRIYSDIYDDDKVYKCFLYNT
                IFTKNPLSRKHLRLNIDLDTPTSILNTVNQYDLVWKIHDSFSKLIQLKNLPQREIPHA
                LRLSISLNCFLDSTTSTSGPVFQRDTVEALRRLTEIWRTYQDWSPPTRGLKHLRDILS
                VLAMYDNPKNLINDVRRT"
CDS             14039. .14713
                /note="YM8520.10, putative DNAJ gene, len: 224, CAI: 0.14,
                similar to DNAJ proteins"
                /codon_start=1
                /product="unknown"
                /protein_id="CAA89797.1"
                /db_xref="GI:825566"
                /db_xref="SWISS-PROT:P48353"
                /translation="MSFTEDQEKIALEILSKDKHEFYEILKVDRKATDSEIKKAYRKL
                AIKLHPDKNSHPKAGEAFKVINRAFEVLSNEEKRSIYDRIGRDPDDDRQMPSRGAASGF
                RGSAGGSPMGGGFEDMFFNSRFGGQRAGPPEDIFDFLFNAGGSPFGASPFGPSASTFS
                FGGPGGFRVYTNNRGGSPFMRQQPRSRQQQQQAEENAVNSQLKNMLVLFIIFIVLPMI
                KDYLFS"

  Query Match             64.0%;  Score 16;  DB 8;  Length 41200;
  Best Local Similarity   100.0%;  Pred. No. 93;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctgtt 18
           ||||||||||||||||
Db   25021 TGCTGGATCTGCTGTT 25036


RESULT  45
AX008043/c
LOCUS       AX008043    46340 bp    DNA             PAT       06-SEP-2000
DEFINITION  Sequence 66 from Patent WO9967395.
ACCESSION   AX008043
VERSION     AX008043.1  GI:9995728
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 46340)
  AUTHORS   Perin,J.P., Rieger,F. and Alliel,P.M.
  TITLE     Nucleic sequence and deduced protein sequence family with human
            endogenous retroviral motifs, and their uses
  JOURNAL   Patent: WO 9967395-A 66 29-DEC-1999;
            INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
            FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES             Location/Qualifiers
     source          1. .46340
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

BASE COUNT 16104 a 8737 c 8435 g 13064 t
ORIGIN

Query Match 64.0%; Score 16; DB 6; Length 46340;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4 gctggatctgctgttt 19
           ||||||||||||||||
Db    1254 GCTGGATCTGCTGTTT 1239
```

Search completed: December 26, 2001, 12:50:57
Job time: 8779 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:16:32 ; Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196F-3
Perfect score: 25
Sequence: 1 gttgctggatctgctgtttgaagcg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 58608

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	100.0	839	22	AAI62941	Human genomic DNA
	2	25	100.0	1041	20	AAX01303	E. coli biotin syn
	3	25	100.0	1084	10	AAN91329	E.coli Bio B gene.
	4	25	100.0	1121	7	AAN60496	Sequence encoding
	5	25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
	6	16	64.0	5000	21	AAZ36139	Nucleotide sequenc
c	7	15	60.0	1683	16	AAT14850	Maize U14.4 DNA se
c	8	15	60.0	1683	16	AAT08746	Maize U14.4 DNA se
c	9	15	60.0	1689	20	AAZ23024	Rat kd312 polypept
c	10	15	60.0	2022	21	AAC44132	Zea mays DNA fragm
c	11	15	60.0	2034	20	AAX04606	Maize genomic clon
c	12	15	60.0	3079	20	AAZ23025	Rat kd312 genomic
	13	15	60.0	6703	19	AAV49536	Adenylate cyclase
	14	14	56.0	275	20	AAZ12718	Human gene express
c	15	14	56.0	336	22	AAF65487	Novel human polynu
c	16	14	56.0	379	20	AAV87319	EST clone BP797.
c	17	14	56.0	383	21	AAC26529	Human secreted pro
c	18	14	56.0	384	18	AAV01888	Human OTK27 gene.
	19	14	56.0	393	22	AAH10789	Human cDNA clone (
	20	14	56.0	464	22	AAI10569	Probe #502 for gen
	21	14	56.0	464	22	AAI31822	Probe #508 used to
	22	14	56.0	464	22	AAI00501	Probe #492 used to
	23	14	56.0	469	22	AAI12900	Probe #2833 for ge
	24	14	56.0	469	22	AAI34258	Probe #2944 used t
	25	14	56.0	469	22	AAI02819	Probe #2810 used t
	26	14	56.0	476	22	AAI22139	Probe #12072 for g
	27	14	56.0	476	22	AAI47435	Probe #16121 used
	28	14	56.0	476	22	AAI07839	Probe #7830 used t
	29	14	56.0	558	22	AAI19821	Probe #9754 for ge
	30	14	56.0	558	22	AAI45016	Probe #13702 used
	31	14	56.0	558	22	AAI05537	Probe #5528 used t
	32	14	56.0	575	22	AAI18864	Probe #8797 for ge
	33	14	56.0	575	22	AAI43986	Probe #12672 used
	34	14	56.0	612	21	AAC99862	Human secreted pro
c	35	14	56.0	619	21	AAA44124	Human secreted exp
c	36	14	56.0	707	21	AAF14018	Aspergillus oryzae
	37	14	56.0	735	16	AAQ92502	Mouse antibody F4-
c	38	14	56.0	753	20	AAZ00347	Nucleotide sequenc
c	39	14	56.0	753	22	AAF89005	Human FATP3 coding
c	40	14	56.0	769	21	AAF15012	Trichoderma reesei
	41	14	56.0	891	20	AAZ06233	Human secreted pro
c	42	14	56.0	1044	21	AAX82834	L-asparaginase DNA
	43	14	56.0	1155	20	AAZ09784	E. coli S-adenosyl
	44	14	56.0	1155	22	AAA99433	Met K oligonucleot
c	45	14	56.0	1173	20	AAZ00358	Nucleotide sequenc
c	46	14	56.0	1173	22	AAF89020	Human FATP3 partia
	47	14	56.0	1220	21	AAC44861	Arabidopsis thalia
c	48	14	56.0	1367	21	AAA81673	N. meningitidis pa
c	49	14	56.0	1493	18	AAV01889	Human OTK27 gene.
c	50	14	56.0	1676	22	AAD03029	Flax 2S storage pr
c	51	14	56.0	1699	22	AAI58178	Human polynucleoti
c	52	14	56.0	1848	18	AAT96346	Chimeric gene cont
c	53	14	56.0	1998	20	AAZ00367	Nucleotide sequenc
c	54	14	56.0	1998	22	AAF89029	Murine FATP3 parti
	55	14	56.0	2080	20	AAZ30162	cDNA encoding a po
c	56	14	56.0	2087	20	AAZ00343	Nucleotide sequenc
c	57	14	56.0	2087	22	AAF89001	Murine FATP3 codin
c	58	14	56.0	2166	20	AAZ00379	Nucleotide sequenc
c	59	14	56.0	2166	22	AAF89041	Human FATP3 coding
c	60	14	56.0	2314	22	AAH17621	Human cDNA sequenc
c	61	14	56.0	2319	22	AAH19223	Human secreted pro
c	62	14	56.0	2339	22	AAD07655	Human secreted pro
c	63	14	56.0	2402	22	AAF89054	Human FATP3 coding
c	64	14	56.0	2405	22	AAF27417	Human fatty acid t
c	65	14	56.0	2447	22	AAH19186	Human secreted pro
c	66	14	56.0	2574	20	AAZ33977	Human PRO703 nucle
c	67	14	56.0	2574	21	AAC78481	Human PRO703 (UNQ3
c	68	14	56.0	2574	21	AAC58239	Human PRO703 nucle
c	69	14	56.0	2669	20	AAZ25020	Human H-beta 58 fa
c	70	14	56.0	2675	21	AAC98021	Human colon cancer
c	71	14	56.0	2709	22	AAH33375	Human colon cancer
c	72	14	56.0	2826	22	AAC85785	cDNA encoding meta
c	73	14	56.0	2925	22	AAC85786	cDNA encoding chim
c	74	14	56.0	3144	22	AAH98356	Human EST-derived
c	75	14	56.0	3243	11	AAQ05686	Islets of Langerha
c	76	14	56.0	3243	20	AAZ32338	Human pancreatic i
c	77	14	56.0	3243	20	AAZ09919	Human islet cell a
c	78	14	56.0	3243	20	AAV63557	Islet cell antibod
c	79	14	56.0	3243	21	AAZ46456	Pancreatic islet c
c	80	14	56.0	3282	16	AAQ80421	Human mGluR5c cDNA
c	81	14	56.0	3421	22	AAH14567	Human cDNA sequenc
c	82	14	56.0	3591	20	AAZ10801	Mouse metabotropic
	83	14	56.0	3720	20	AAZ09790	Plasmid pHS1 metK
c	84	14	56.0	3918	20	AAZ10803	Rat metabotropic g

c	85	14	56.0	4085	16	AAQ80419	Human mGluR5a cDNA
c	86	14	56.0	4181	16	AAQ80420	Human mGluR5b cDNA
c	87	14	56.0	4207	22	AAD04990	Human metabotropic
c	88	14	56.0	4207	22	AAD05029	Human metabotropic
c	89	14	56.0	4207	22	AAD05030	Human metabotropic
c	90	14	56.0	4207	22	AAD05033	Human metabotropic
c	91	14	56.0	4303	22	AAD04991	Human metabotropic
c	92	14	56.0	4303	22	AAD05031	Human metabotropic
c	93	14	56.0	4303	22	AAD05032	Human metabotropic
c	94	14	56.0	4566	19	AAV41550	Nucleotide sequenc
c	95	14	56.0	4566	20	AAX01089	Human G-protein co
c	96	14	56.0	4568	18	AAT44039	Human G-protein re
c	97	14	56.0	4568	20	AAX01094	Human G-protein co
	98	14	56.0	4975	20	AAZ09792	Plasmid pHS1 metK/
	99	14	56.0	5319	19	AAV35235	R. prowazekii S-la
c	100	14	56.0	7326	16	AAQ99456	Sequence encoding

ALIGNMENTS

RESULT 1
AAI62941
ID AAI62941 standard; DNA; 839 BP.
XX
AC AAI62941;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human genomic DNA SEQ ID NO 269.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200155449-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01346.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0218290.
PR 14-AUG-2000; 2000US-0225447.
PR 01-SEP-2000; 2000US-0229343.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231243.
PR 25-SEP-2000; 2000US-0234997.
PR 29-SEP-2000; 2000US-0236367.
PR 13-OCT-2000; 2000US-0239937.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246528.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249265.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-476225/51.
XX
PT Novel plasma membrane associated proteins useful for diagnosing,
PT treating, preventing and/or prognosing disorders related to the
PT proteins, including cancer, immune response and neuronal disorders -
XX
PS Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;

Query Match 100.0%; Score 25; DB 22; Length 839;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db      89 gttgctggatctgctgtttgaagcg 113
```

RESULT 2
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
PR 08-MAR-1995; 95US-0401068.
XX

```
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Patton DA;
XX
DR   WPI; 1999-152902/13.
DR   P-PSDB; AAW73906.
XX
PT   Transgenic plants with high biotin levels - transformed with DNA
PT   encoding di:amino-pelargonic acid amino-transferase or biotin
PT   synthase
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.00017;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db     57 gttgctggatctgctgtttgaagcg 81


RESULT   3
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.00017;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db     80 gttgctggatctgctgtttgaagcg 104


RESULT   4
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 0.00017;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtttgaagcg 25
          |||||||||||||||||||||||||
Db     98 gttgctggatctgctgtttgaagcg 122


RESULT   5
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
```

```
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag=  a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag=  b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag=  c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag=  d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag=  e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag=  f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag=  g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag=  h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag=  i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag=  j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag=  k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag=  l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag=  m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag=  n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                    7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag=  o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag=  p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag=  q
FT                   /standard_name= "rho-independent transcriptional
FT                                    terminator"
FT   stem_loop       5583..5605
FT                   /*tag=  r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;   93WO-EP02688.
XX
PR   02-OCT-1992;   92CH-0003124.
PR   15-JUL-1993;   93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.00017;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db     173 gttgctggatctgctgtttgaagcg 197


RESULT   6
AAZ36139
ID   AAZ36139 standard; DNA; 5000 BP.
XX
AC   AAZ36139;
```

```
XX
DT   11-FEB-2000  (first entry)
XX
DE   Nucleotide sequence of the DDM1 gene which regulates DNA methylation.
XX
KW   DDM1 gene; chromosome 5; zinc-finger protein gene; SWI2/SNF2 family;
KW   glutamic acid tRNA gene; genomic cytosine methylation; gene silencing;
KW   DNA methylation; chromatin structure; DNA hypomethylation;
KW   DDM1 gene allele; inbreeding depression; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             535..4354
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            535..566
FT                   /*tag= b
FT                   /number= 1
FT   intron          567..771
FT                   /*tag= c
FT                   /number= 1
FT   exon            772..850
FT                   /*tag= d
FT                   /number= 2
FT   allele          replace(785, (N82))
FT                   /*tag= e
FT                   /note= "site of the ddm1-5 (som8) mutation; this base is
FT                         replaced with 82 unspecified bases"
FT   intron          851..985
FT                   /*tag= f
FT                   /number= 2
FT   exon            986..1252
FT                   /*tag= g
FT                   /number= 3
FT   intron          1253..1353
FT                   /*tag= h
FT                   /number= 3
FT   exon            1354..1440
FT                   /*tag= i
FT                   /number= 4
FT   intron          1441..1548
FT                   /*tag= j
FT                   /number= 4
FT   exon            1549..1895
FT                   /*tag= k
FT                   /number= 5
FT   intron          1896..1975
FT                   /*tag= l
FT                   /number= 5
FT   exon            1976..2165
FT                   /*tag= m
FT                   /number= 6
FT   intron          2164..2250
FT                   /*tag= n
FT                   /number= 6
FT   exon            2251..2426
FT                   /*tag= o
FT                   /number= 7
FT   allele          replace(2384..2385)
FT                   /*tag= p
FT                   /note= "site of the ddm1-6 mutation; either the G
FT                         at 2384 or 2385 is deleted"
FT   intron          2427..2558
FT                   /*tag= q
FT                   /number= 7
FT   exon            2559..2625
FT                   /*tag= r
FT                   /number= 8
FT   intron          2626..2702
FT                   /*tag= s
FT                   /number= 8
FT   exon            2703..2892
FT                   /*tag= t
FT                   /number= 9
FT   intron          2893..2974
FT                   /*tag= u
FT                   /number= 9
FT   exon            2975..3070
FT                   /*tag= v
FT                   /number= 10
FT   intron          3071..3147
FT                   /*tag= w
FT                   /number= 10
FT   exon            3148..3242
FT                   /*tag= x
FT                   /number= 11
FT   misc_feature    3186
FT                   /*tag= y
FT                   /note= "alternative splice donor site used in ddm1-2"
FT   intron          3243..3316
FT                   /*tag= z
FT                   /number= 11
FT   allele          replace(3243, A)
FT                   /*tag= aa
FT                   /note= "site of the ddm1-2 mutation"
FT   exon            3317..3436
FT                   /*tag= ab
FT                   /number= 12
FT   allele          replace(3337, A)
FT                   /*tag= ac
FT                   /note= "site of the ddm1-7 (som5) mutation"
FT   intron          3437..3539
FT                   /*tag= ad
FT                   /number= 12
FT   exon            3540..3659
FT                   /*tag= ae
FT                   /number= 13
FT   intron          3660..3744
FT                   /*tag= af
FT                   /number= 13
FT   exon            3745..3843
FT                   /*tag= ag
FT                   /number= 14
FT   intron          3844..3933
FT                   /*tag= ah
FT                   /number= 14
FT   exon            3934..4038
FT                   /*tag= ai
FT                   /number= 15
FT   intron          4039..4129
FT                   /*tag= aj
FT                   /number= 15
FT   exon            4130..4354
FT                   /*tag= ak
FT                   /number= 16
FT   tRNA            4755..4826
FT                   /*tag= al
FT                   /note= "complement of predicted tRNA-glu"
XX
PN   WO9955891-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09268.
XX
PR   24-JUN-1998;   98US-0104070.
XX
PA   (UNIW ) UNIV WASHINGTON.
XX
PI   Richards EJ,  Jeddeloh JA;
XX
DR   WPI; 2000-013439/01.
DR   P-PSDB; AAY43847.
XX
PT   Arabidopsis thaliana DDM1 gene alters methylation state of DNA, useful
```

```
PT   for reducing gene silencing or inbreeding depression -
XX
PS   Claim 6; Page 36-38; 59pp; English.
XX
CC   The present sequence represents the DDM1 gene of Arabidopsis thaliana.
CC   The gene is located on chromosome 5, lower arm, where it is flanked on
CC   the centromeric side, within 20 kb, by a zinc-finger protein gene and
CC   on the telomeric side, within 1 kb, by a glutamic acid tRNA gene. The
CC   DDM1 protein is a member of the SWI2/SNF2 family of polypeptides. The
CC   DDM1 protein is important for the maintenance of proper genomic cytosine
CC   methylation, and its function appears to be necessary to maintain gene
CC   silencing. The protein appears to function in the DNA methylation system
CC   by affecting chromatin structure. Disruption of the gene is associated
CC   with DNA hypomethylation. Four DDM1 gene alleles have been identified,
CC   all of which are expected to destroy or severely reduce gene function.
CC   The DDM1 nucleic acid can be used to create transgenic plants with
CC   altered levels of DNA methylation. The mutant or engineered plants are
CC   expected to be viable and display valuable characteristic such as
CC   inbreeding depression or a reduction in gene silencing.
XX
SQ   Sequence 5000 BP; 1386 A; 909 C; 1029 G; 1676 T; 0 other;


  Query Match             64.0%;  Score 16;  DB 21;  Length 5000;
  Best Local Similarity   100.0%;  Pred. No. 10;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctgtt 18
          ||||||||||||||||
Db   1132 tgctggatctgctgtt 1147


RESULT   7
AAT14850/c
ID   AAT14850 standard; DNA; 1683 BP.
XX
AC   AAT14850;
XX
DT   04-JUL-1996  (first entry)
XX
DE   Maize U14.4 DNA sequence.
XX
KW   snoRNA; small nucleolar RNA; U14; non-translated sequence; conserved;
KW   Box C; Box D; processing; pre-ribosomal RNA; ribosome formation;
KW   normal growth phenotype; yeast; plant; maize; potato; 18S rRNA; primer;
KW   PCR; polymerase chain reaction; amplification; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    379..385
FT                   /*tag= a
FT                   /note= "nucleotides not identified in sequence"
FT   repeat_unit     1212..1217
FT                   /*tag= b
FT                   /note= "inverted repeat"
FT   misc_RNA        1218..1329
FT                   /*tag= c
FT                   /note= "MU14.4"
FT   misc_feature    1220..1225
FT                   /*tag= d
FT                   /note= "Box C"
FT   misc_feature    1324..1329
FT                   /*tag= e
FT                   /note= "Box D"
FT   repeat_unit     1330..1335
FT                   /*tag= f
FT                   /note= "inverted repeat"
XX
PN   WO9530748-A2.
XX
PD   16-NOV-1995.
XX
PF   04-MAY-1995;   95WO-EP01694.
XX
PR   09-DEC-1994;   94EP-0119487.
PR   04-MAY-1994;   94WO-EP01409.
XX
PA   (GENE-) GENE SHEARS PTY LTD.
XX
PI   Brown JWS,  Leader DJ,  Waugh R;
XX
DR   WPI; 1995-404115/51.
XX
PT   Plant U14 nucleic acid sequences and derivatives - regulate
PT   ribosomal RNA production and accumulation in plants
XX
PS   Example; Fig 7A-D; 57pp; English.
XX
CC   The present sequence shows the maize single U14 sequence U14.4,
CC   including flanking intergenic sequences and inverted repeats, as obtd.
CC   by sequencing of fragments of isolated genomic clones. Inverted repeats
CC   and Box C and D elements are shown above (see AAT08751-52). U14 genes
CC   are small nucleolar RNAs present in the genome as non-translated
CC   sequences. SnoRNAs are involved in processing of pre-ribosomal RNA
CC   transcripts and ribosome formation. U14snoRNA is required for a normal
CC   growth phenotype and for processing of pre-rRNA transcripts in yeast.
CC   Plant U14 sequences have been identified and characterised in the present
CC   invention, and may be intron encoded.
XX
SQ   Sequence 1683 BP; 383 A; 387 C; 358 G; 548 T; 7 other;


  Query Match             60.0%;  Score 15;  DB 16;  Length 1683;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgct 15
          |||||||||||||||
Db    601 GTTGCTGGATCTGCT 587


RESULT   8
AAT08746/c
ID   AAT08746 standard; DNA; 1683 BP.
XX
AC   AAT08746;
XX
DT   04-JUL-1996  (first entry)
XX
DE   Maize U14.4 DNA sequence.
XX
KW   snoRNA; small nucleolar RNA; U14; non-translated sequence; conserved;
KW   Box C; Box D; processing; pre-ribosomal RNA; ribosome formation;
KW   normal growth phenotype; yeast; plant; maize; potato; 18S rRNA; primer;
KW   PCR; polymerase chain reaction; amplification; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    379..385
FT                   /*tag= a
FT                   /note= "nucleotides not identified in sequence"
FT   repeat_unit     1212..1217
FT                   /*tag= b
FT                   /note= "inverted repeat"
FT   misc_RNA        1218..1329
FT                   /*tag= c
FT                   /note= "MU14.4"
FT   misc_feature    1220..1225
FT                   /*tag= d
FT                   /note= "Box C"
FT   misc_feature    1324..1329
FT                   /*tag= e
FT                   /note= "Box D"
FT   repeat_unit     1330..1335
```

```
FT                  /*tag=  f
FT                  /note= "inverted repeat"
XX
PN   WO9530747-A1.
XX
PD   16-NOV-1995.
XX
PF   04-MAY-1994;   94WO-EP01409.
XX
PR   04-MAY-1994;   94WO-EP01409.
XX
PA   (GENE-) GENE SHEARS PTY LTD.
XX
PI   Brown JWS,  Leader DJ,  Waugh R;
XX
DR   WPI; 1995-404114/51.
XX
PT   Nucleic acid contg. plant U14 sequences and their derivs. - for
PT   regulating prodn. of ribosomal RNA, also as stabilisers for
PT   heterologous RNA
XX
PS   Example; Fig 7A-D; 57pp; English.
XX
CC   The present sequence shows the maize single U14 sequence U14.4,
CC   including flanking intergenic sequences and inverted repeats, as obtd.
CC   by sequencing of fragments of isolated genomic clones. Inverted repeats
CC   and Box C and D elements are shown above (see AAT08729-30). U14 genes
CC   are small nucleolar RNAs present in the genome as non-translated
CC   sequences. SnoRNAs are involved in processing of pre-ribosomal RNA
CC   transcripts and ribosome formation. U14snoRNA is required for a normal
CC   growth phenotype and for processing of pre-rRNA transcripts in yeast.
CC   Plant U14 sequences have been identified and characterised in the present
CC   invention, and may be intron encoded.
XX
SQ   Sequence 1683 BP; 383 A; 387 C; 358 G; 548 T; 7 other;


  Query Match             60.0%;  Score 15;  DB 16;  Length 1683;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgct 15
          |||||||||||||||
Db    601 GTTGCTGGATCTGCT 587


RESULT   9
AAZ23024/c
ID   AAZ23024 standard; cDNA; 1689 BP.
XX
AC   AAZ23024;
XX
DT   17-JAN-2000  (first entry)
XX
DE   Rat kd312 polypeptide encoding cDNA.
XX
KW   kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW   heart attack; head trauma; neurodegenerative disease; rat;
KW   Parkinson's disease; Alzheimer's disease; ss.
XX
OS   Rattus sp.
XX
PN   WO9950288-A2.
XX
PD   07-OCT-1999.
XX
PF   30-MAR-1999;   99WO-US06993.
XX
PR   31-MAR-1998;   98US-0053374.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Yen K;
XX
DR   WPI; 1999-601322/51.
DR   P-PSDB; AAY42694.
XX
PT   kd312 polypeptides useful for treating diseases and disorders
PT   associated with alterations in cell proliferation and cell death  -
XX
PS   Claim 2; Fig 9; 85pp; English.
XX
CC   The invention provides nucleic acid molecules encoding human and rat
CC   kd312 polypeptides. The kd312 polypeptides can be expressed by standard
CC   recombinant methodology. The kd312 sequences, and the antibodies against
CC   the proteins may be used to treat or diagnose the presence or progression
CC   of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
CC   stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
CC   Parkinson's disease and Alzheimer's disease). The present sequence
CC   represents the rat kd312 cDNA sequence.
XX
SQ   Sequence 1689 BP; 417 A; 462 C; 434 G; 376 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 20;  Length 1689;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      7 ggatctgctgtttga 21
          |||||||||||||||
Db    495 GGATCTGCTGTTTGA 481


RESULT  10
AAC44132/c
ID   AAC44132 standard; DNA; 2022 BP.
XX
AC   AAC44132;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Zea mays DNA fragment SEQ ID NO: 41750.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic;
KW   pathway; promoter; termination sequence; corn; ss.
XX
OS   Zea mays subsp. mays.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
```

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

```
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             60.0%;  Score 15;  DB 21;  Length 2022;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaagc 24
           |||||||||||||||
Db    1268 TCTGCTGTTTGAAGC 1254


RESULT  11
AAX04606/c
ID   AAX04606 standard; DNA; 2034 BP.
XX
AC   AAX04606;
XX
DT   07-APR-1999  (first entry)
XX
DE   Maize genomic clone MzU14.4 snoR cluster sequence.
XX
KW   Small nucleolar RNA gene; snoR; snoR1.4; maize; promoter;
KW   RNA production; RNA stability; small non-translated RNA;
KW   intergenic region; U49.1; SnoR2.4; U14.4; SnoR3.4; ds.
XX
OS   Zea sp.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..890
FT                   /*tag=  a
FT                   /note= "upstream region having promoter activity"
FT   misc_feature    891..983
FT                   /*tag=  b
FT                   /note= "SnoR1.4 coding sequence; no protein given in
FT                          specification"
FT   misc_feature    984..1100
FT                   /*tag=  c
FT                   /note= "intergenic region"
FT   misc_feature    1101..1294
FT                   /*tag=  d
FT                   /note= "U49.1 coding sequence; no protein given in
FT                          specification"
FT   misc_feature    1295..1361
FT                   /*tag=  e
FT                   /note= "intergenic region"
FT   misc_feature    1362..1512
FT                   /*tag=  f
FT                   /note= "SnoR2.4 coding sequence; no protein given in
FT                          specification"
FT   misc_feature    1513..1587
FT                   /*tag=  g
FT                   /note= "intergenic region"
FT   misc_feature    1588..1711
FT                   /*tag=  h
FT                   /note= "U14.4 coding sequence; no protein given in
FT                          specification"
FT   misc_feature    1712..1743
FT                   /*tag=  i
FT                   /note= "intergenic region"
FT   misc_feature    1744..1873
FT                   /*tag=  j
FT                   /note= "SnoR3.1 coding sequence; no protein given in
FT                          specification"
FT   misc_feature    1874..2034
FT                   /*tag=  k
FT                   /note= "3' flanking sequence"
XX
PN   EP887405-A1.
XX
PD   30-DEC-1998.
XX
PF   25-JUN-1997;   97EP-0401480.
XX
PR   25-JUN-1997;   97EP-0401480.
XX
PA   (GENE-) GENE SHEARS PTY LTD.
XX
PI   Brown JWS,  Leader DJ;
XX
DR   WPI; 1999-047872/05.
XX
PT   Producing stabilised RNA molecules  - using a DNA precursor
PT   comprising plant small nucleolar RNA stabilising sequences
XX
PS   Claim 46; Fig 18a-f; 98pp; English.
XX
CC   The present sequence represents the full sequence of the the small
CC   nucleolar RNA (snoRNA) cluster of maize genomic clone MzU14.4. The
CC   SnoR genes in the cluster are SnoR1.4, U49.1, SnoR2.4, U14.4, and
CC   SnoR3.1. Plant snoR genes enable production of multiple stabilised
CC   RNA from a single promoter in a splicing-independent manner,
CC   and allow the production of RNA from intron or non-intron sequences.
CC   The present sequence is used in the method of the invention. The
CC   specification describes a method for producing small non-translated
CC   RNA molecules, which comprises introducing a precursor DNA containing
CC   one or a cluster of coding sequences into a plant or yeast cell. Each
CC   coding sequence comprises a non-translated RNA sequence to be stabilised,
CC   which is capable of functionally interacting with a cellular component
CC   and stabilising sequences comprising at least one protein binding site.
CC   The stabilised RNA units are transcribed by the cell's endogenous
CC   processing machinery. The process is useful for qualitatively and
CC   quantitatively enhancing the production and stability of RNA molecules,
CC   especially non-translated RNA e.g ribozymes and antisense RNA. The
CC   nucleic acid molecules are useful as probes or primers for identification
CC   of snoRNA sequences and intergenic regions.
XX
SQ   Sequence 2034 BP; 449 A; 504 C; 435 G; 646 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 20;  Length 2034;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgct 15
           |||||||||||||||
Db     971 GTTGCTGGATCTGCT 957


RESULT  12
AAZ23025/c
ID   AAZ23025 standard; DNA; 3079 BP.
XX
AC   AAZ23025;
XX
DT   17-JAN-2000  (first entry)
XX
DE   Rat kd312 genomic DNA sequence.
XX
KW   kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW   heart attack; head trauma; neurodegenerative disease; rat;
KW   Parkinson's disease; Alzheimer's disease; ss.
XX
OS   Rattus sp.
XX
```

PN WO9950288-A2.
XX
PD 07-OCT-1999.
XX
PF 30-MAR-1999; 99WO-US06993.
XX
PR 31-MAR-1998; 98US-0053374.
XX
PA (AMGE-) AMGEN INC.
XX
PI Yen K;
XX
DR WPI; 1999-601322/51.
DR P-PSDB; AAY42694.
XX
PT kd312 polypeptides useful for treating diseases and disorders
PT associated with alterations in cell proliferation and cell death -
XX
PS Claim 2; Fig 7; 85pp; English.
XX
CC The invention provides nucleic acid molecules encoding human and rat
CC kd312 polypeptides. The kd312 polypeptides can be expressed by standard
CC recombinant methodology. The kd312 sequences, and the antibodies against
CC the proteins may be used to treat or diagnose the presence or progression
CC of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
CC stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
CC Parkinson's disease and Alzheimer's disease). The present sequence
CC represents the rat kd312 genomic DNA sequence.
XX
SQ Sequence 3079 BP; 708 A; 843 C; 829 G; 699 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 3079;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ggatctgctgtttga 21
||||||||||||||||
Db 803 GGATCTGCTGTTTGA 789

RESULT 13
AAV49536
ID AAV49536 standard; DNA; 6703 BP.
XX
AC AAV49536;
XX
DT 16-NOV-1998 (first entry)
XX
DE Adenylate cyclase gene.
XX
KW ds; adenylate cyclase; recombinant microorganism; chemical production;
KW riboflavin; carotenoid.
XX
OS Ashbya gossypii.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..670
FT /*tag= a
FT CDS 671..6295
FT /*tag= b
FT /product= "Adenylate cyclase"
FT 3'UTR 6296..6703
FT /*tag= c
XX
PN WO9829538-A2.
XX
PD 09-JUL-1998.
XX
PF 29-DEC-1997; 97WO-EP07309.
XX
PR 31-DEC-1996; 97CH-0000016.
XX
PA (BADI) BASF AG.
XX
PI Althoefer H, Altmann-Joehl R, Philippsen P, Seulberger H;
XX
DR WPI; 1998-388120/33.
DR P-PSDB; AAW64518.
XX
PT New gene for adenylate cyclase from Ashbya gossypii - useful for
PT generating recombinant microorganisms with alteration in gene of
PT cAMP-dependent signalling pathway for increasing production of fine
PT chemicals
XX
PS Claim 1; Page 10-20; 33pp; German.
XX
CC The adenylate cyclase (AC) gene is used to generate recombinant
CC microorganisms for production of chemicals, particularly riboflavin but
CC also others useful in human or animal nutrition, cosmetics and medicines,
CC and also carotenoids for use as natural colourings. The modified
CC microorganisms produce chemicals at higher levels than wild-type
CC organisms. Particularly, for riboflavin, cells with a disrupted AC gene
CC are grown on glucose-containing medium.
XX
SQ Sequence 6703 BP; 1954 A; 1527 C; 1541 G; 1681 T; 0 other;

Query Match 60.0%; Score 15; DB 19; Length 6703;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgctggatctgctgt 17
|||||||||||||||
Db 6234 tgctggatctgctgt 6248

RESULT 14
AAZ12718
ID AAZ12718 standard; cDNA; 275 BP.
XX
AC AAZ12718;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:187.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX

```
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 705; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 275 BP; 65 A; 44 C; 48 G; 66 T; 52 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 275;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgct 15
           ||||||||||||||
Db     169 ttgctggatctgct 182


RESULT  15
AAF65487/c
ID   AAF65487 standard; cDNA; 336 BP.
XX
AC   AAF65487;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 1243.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
PI   Reinhard C,  Randazzo F,  Kennedy GC,  Pot D,  Lamson G,  Drmanac R;
PI   Crkenjakov R,  Drmanac S,  Dickson M,  Labat I,  Leshkowitiz D;
PI   Kita D,  Garcia V,  Jones LW,  Strache-Crain B;
XX
DR   WPI; 2001-091805/10.
XX
PT   Library of polynucleotides for diagnosing a cancerous state of a
PT   mammalian cell and detecting cancer, particularly of the colon or
PT   prostate, comprises 3351 human polynucleotide sequences -
XX
PS   Claim 9; Page 722; 1046pp; English.
XX
CC   The present sequence is one of 3351 sequences in a library of human
CC   polynucleotides. The library is used to detect differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell and can
CC   detect colon, prostate, breast and lung cancer. The library can be used
CC   to produce probes for detection of mRNA and to produce additional copies
CC   of the polynucleotides. The probes can be used for chromosome mapping of
CC   the polynucleotide and for detection of transcription levels. Ribozymes
CC   or antisense oligonucleotides can be generated. The polynucleotides and
CC   their gene products are used as genetic or biochemical markers (e.g. in
CC   blood or tissues) that will detect the earliest changes along the
CC   carcinogenesis pathway and/or monitor the efficacy of therapies and
CC   preventive interventions. The polynucleotides, polypeptides and
CC   antibodies against them can be used in pharmaceutical compositions to
CC   treat the cancers and proliferative disorders such as neoplasia,
CC   dysplasia and hyperplasia.
XX
SQ   Sequence 336 BP; 97 A; 84 C; 59 G; 96 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 336;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 atctgctgtttgaa 22
           ||||||||||||||
Db     159 ATCTGCTGTTTGAA 146


RESULT  16
AAV87319/c
ID   AAV87319 standard; cDNA; 379 BP.
XX
AC   AAV87319;
XX
DT   27-APR-1999  (first entry)
XX
DE   EST clone BP797.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845435-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06954.
XX
PR   10-APR-1997;   97US-0835913.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070076/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 527; 633pp; English.
XX
CC   This sequence represents an expressed sequence tag (EST), and is a
```

```
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 379 BP; 124 A; 68 C; 70 G; 116 T; 1 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 379;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 gatctgctgtttga 21
           ||||||||||||||
Db     129 GATCTGCTGTTTGA 116


RESULT  17
AAC26529/c
ID   AAC26529 standard; cDNA; 383 BP.
XX
AC   AAC26529;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 30604.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 30604; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 383 BP; 96 A; 103 C; 63 G; 119 T; 2 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 383;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgc 14
           ||||||||||||||
Db     263 GTTGCTGGATCTGC 250


RESULT  18
AAV01888/c
ID   AAV01888 standard; cDNA; 384 BP.
XX
AC   AAV01888;
XX
DT   20-APR-1998  (first entry)
XX
DE   Human OTK27 gene.
XX
KW   Human; foetal brain cDNA library; GDP dissociation stimulating protein;
KW   brain specific nucleosome assembly protein; diagnosis; therapy;
KW   skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NPIK;
KW   nel-related protein type 1; nel-related type 2; hereditary disease;
KW   cancer; OTK27; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..384
FT                   /*tag= a
FT                   /product= "OTK27"
FT                   /note= "no stop codon given"
XX
PN   EP796913-A2.
XX
PD   24-SEP-1997.
XX
PF   19-MAR-1997;   97EP-0104842.
XX
PR   05-MAR-1997;   97JP-0069163.
PR   19-MAR-1996;   96JP-0063410.
XX
PA   (SAKA ) OTSUKA PHARM CO LTD.
XX
PI   Fujiwara T,  Horie M,  Watanabe T;
XX
DR   WPI; 1997-459830/43.
DR   P-PSDB; AAW37503.
XX
PT   Novel human genes, e.g. brain-specific nucleosome assembly protein -
PT   useful for diagnosis or therapy of hereditary disease and cancer
XX
PS   Example 3; Page 35; 123pp; English.
XX
CC   The present sequence encodes a OTK27 isolated from a human foetal brain
CC   cDNA library. The nucleotide or amino acid sequences are useful for
CC   in-vitro diagnosis of hereditary diseases and cancer and for preparation
CC   of pharmaceuticals.
XX
SQ   Sequence 384 BP; 94 A; 111 C; 98 G; 81 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 18;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tggatctgctgttt 19
```

```
          ||||||||||||||
Db    350 TGGATCTGCTGTTT 337


RESULT  19
AAH10789
ID   AAH10789 standard; cDNA; 393 BP.
XX
AC   AAH10789;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:7624.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 3; SEQ ID 7624; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 393 BP; 89 A; 100 C; 115 G; 82 T; 7 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 393;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db     293 tctgctgtttgaag 306


RESULT  20
AAI10569
ID   AAI10569 standard; DNA; 464 BP.
XX
AC   AAI10569;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #502 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 502; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 464 BP; 77 A; 137 C; 140 G; 110 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gctggatctgctgt 17
           ||||||||||||||
Db     165 gctggatctgctgt 178


RESULT  21
AAI31822
ID   AAI31822 standard; DNA; 464 BP.
XX
```

```
AC   AAI31822;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #508 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 508; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 464 BP; 77 A; 137 C; 140 G; 110 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gctggatctgctgt 17
           ||||||||||||||
Db     165 gctggatctgctgt 178


RESULT  22
AAI00501
ID   AAI00501 standard; DNA; 464 BP.
XX
AC   AAI00501;
XX
DT   09-OCT-2001  (first entry)
XX                                                          !
DE   Probe #492 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 492; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 464 BP; 77 A; 137 C; 140 G; 110 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 464;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gctggatctgctgt 17
           ||||||||||||||
Db     165 gctggatctgctgt 178


RESULT  23
AAI12900
ID   AAI12900 standard; DNA; 469 BP.
XX
AC   AAI12900;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #2833 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
```

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID No 2833; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 469 BP; 81 A; 90 C; 145 G; 153 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 469;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 gatctgctgtttga 21
||||||||||||||
Db 295 gatctgctgtttga 308

RESULT 24
AAI34258
ID AAI34258 standard; DNA; 469 BP.
XX
AC AAI34258;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #2944 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 2944; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 469 BP; 81 A; 90 C; 145 G; 153 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 469;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 gatctgctgtttga 21
||||||||||||||
Db 295 gatctgctgtttga 308

RESULT 25
AAI02819
ID AAI02819 standard; DNA; 469 BP.
XX
AC AAI02819;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #2810 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID No 2810; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed

```
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 469 BP; 81 A; 90 C; 145 G; 153 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 469;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 gatctgctgtttga 21
          ||||||||||||||
Db    295 gatctgctgtttga 308


RESULT  26
AAI22139
ID   AAI22139 standard; DNA; 476 BP.
XX
AC   AAI22139;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #12072 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 12072; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 476;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 gatctgctgtttga 21
          ||||||||||||||
Db    135 gatctgctgtttga 148


RESULT  27
AAI47435
ID   AAI47435 standard; DNA; 476 BP.
XX
AC   AAI47435;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #16121 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 16121; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 476;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 gatctgctgtttga 21
          ||||||||||||||
Db    135 gatctgctgtttga 148


RESULT  28
AAI07839
ID   AAI07839 standard; DNA; 476 BP.
XX
AC   AAI07839;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Probe #7830 used to measure gene expression in human breast sample.
XX
```

```
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 7830; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 476;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 gatctgctgtttga 21
           ||||||||||||||
Db     135 gatctgctgtttga 148


RESULT  29
AAI19821
ID   AAI19821 standard; DNA; 558 BP.
XX
AC   AAI19821;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #9754 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 9754; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 558 BP; 95 A; 154 C; 169 G; 140 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gctggatctgctgt 17
           ||||||||||||||
Db     105 gctggatctgctgt 118


RESULT  30
AAI45016
ID   AAI45016 standard; DNA; 558 BP.
XX
AC   AAI45016;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #13702 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
```

```
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 13702; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 558 BP; 95 A; 154 C; 169 G; 140 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gctggatctgctgt 17
           ||||||||||||||
Db     105 gctggatctgctgt 118


RESULT  31
AAI05537
ID   AAI05537 standard; DNA; 558 BP.
XX
AC   AAI05537;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Probe #5528 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 5528; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 558 BP; 95 A; 154 C; 169 G; 140 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gctggatctgctgt 17
           ||||||||||||||
Db     105 gctggatctgctgt 118


RESULT  32
AAI18864
ID   AAI18864 standard; DNA; 575 BP.
XX
AC   AAI18864;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #8797 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 8797; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
```

```
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 575 BP; 174 A; 128 C; 161 G; 112 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 575;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 tggatctgctgttt 19
          ||||||||||||||
Db    222 tggatctgctgttt 235


RESULT  33
AAI43986
ID   AAI43986 standard; DNA; 575 BP.
XX
AC   AAI43986;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #12672 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 12672; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 575 BP; 174 A; 128 C; 161 G; 112 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 575;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 tggatctgctgttt 19
          ||||||||||||||
Db    222 tggatctgctgttt 235


RESULT  34
AAC99862
ID   AAC99862 standard; cDNA; 612 BP.
XX
AC   AAC99862;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Human secreted protein gene 45 SEQ ID NO:55.
XX
KW   Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW   cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW   fungicide; ophthalmological; gene therapy; pathological condition;
KW   autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW   neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
KW   cerebrovascular disorder; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; skin aging; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200070042-A1.
XX
PD   23-NOV-2000.
XX
PF   11-MAY-2000; 2000WO-US12788.
XX
PR   13-MAY-1999;   99US-0134068.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Moore PA,  Young PE,  Komatsoulis GA,  Birse CE;
PI   Duan RD,  Florence KA,  Soppet DR;
XX
DR   WPI; 2000-679828/66.
DR   P-PSDB; AAB56121.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 1; Page 881; 1065pp; English.
XX
CC   The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC   human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC   proteins have activities based on the tissues and cells the genes are
CC   expressed in. Examples of activities include: immunosuppressive;
CC   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC   vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC   virucide; fungicide; and opthalmalogical. The human secreted
CC   polynucleotides and proteins can be used to prevent, treat or ameliorate
CC   a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC   dogs, chickens or sheep. They are also used in diagnosing a pathological
CC   condition or susceptibility to a pathological condition. Disorders which
CC   are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC   arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC   liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC   disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC   e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC   fungi and ocular disorders e.g. corneal infection. The proteins can also
CC   be used to aid wound healing and epithelial cell proliferation, to
CC   prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The proteins can also be used as a
CC   food additive or preservative to increase or decrease storage
CC   capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC   in the exemplification of the present invention.
XX
SQ   Sequence 612 BP; 199 A; 139 C; 139 G; 135 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 612;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       2 ttgctggatctgct 15
           ||||||||||||||
Db     372 ttgctggatctgct 385


RESULT  35
AAA44124/c
ID   AAA44124 standard; cDNA; 619 BP.
XX
AC   AAA44124;
XX
DT   21-AUG-2000  (first entry)
XX
DE   Human secreted expressed sequence tag SEQ ID NO:699.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW   vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW   tumour; infection; depression; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200021991-A1.
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;   99WO-US24206.
XX
PR   15-OCT-1998;   98US-0104436.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR;
XX
DR   WPI; 2000-317938/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (sESTs), useful for treating various disorders
PT   such as autoimmune, infectious, and central nervous system disorders -
XX
PS   Claim 1; Page 383; 803pp; English.
XX
CC   AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC   sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC   tissue sources. The sESTs can have a range of activities depending on
CC   the tissues they were isolated from. The activities include:
CC   chemotactic; proliferative; immunomodulatory; haematopoietic;
CC   chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC   cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC   antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC   nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC   anticonvulsant; and antidepressant. The sESTs can be used for gene
CC   therapy and in vaccines. The sESTs are useful as probes for the
CC   identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC   disease), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC   in the exemplification of the present invention.
XX
SQ   Sequence 619 BP; 193 A; 123 C; 125 G; 178 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 619;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 atctgctgtttgaa 22
           ||||||||||||||
Db      43 ATCTGCTGTTTGAA 30


RESULT  36
AAF14018/c
ID   AAF14018 standard; cDNA; 707 BP.
XX
AC   AAF14018;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Aspergillus oryzae EST SEQ ID NO:6541.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM,  Rey MW,  Shuster JR,  Kauppinen S,  Clausen IG,  Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 88; Page 2678; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway
CC   engineering. Using ESTs provides several advantages over genomic or
CC   random cDNA clones including elimination of redundancy as one spot on an
```

```
CC   array equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 707 BP; 154 A; 241 C; 156 G; 156 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 707;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ctggatctgctgtt 18
           ||||||||||||||
Db     693 CTGGATCTGCTGTT 680


RESULT  37
AAQ92502
ID   AAQ92502 standard; cDNA; 735 BP.
XX
AC   AAQ92502;
XX
DT   07-FEB-1996  (first entry)
XX
DE   Mouse antibody F4-7 heavy chain variable region coding sequence.
XX
KW   Primer; amplification; PCR; mouse; kappa chain; heavy chain; Fab;
KW   antibody; immunotolerance; animal; variegated display library;
KW   variable region; antigen; immunorecessive; cell surface marker; foetal;
KW   cancer; stem cell; variant; therapy; Alzheimer's disease; hybridoma;
KW   familial hypercholesterolaemia; binding affinity; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             67..735
FT                   /*tag=  a
FT                   /product= antibody F4-7 heavy chain variable region
XX
PN   WO9515982-A2.
XX
PD   15-JUN-1995.
XX
PF   08-DEC-1994;   94WO-US14106.
XX
PR   06-DEC-1994;   94US-0350400.
PR   08-DEC-1993;   93US-0164022.
XX
PA   (GENZ ) GENZYME CORP.
XX
PI   Barsomian G,  Copeland DP,  Hillhouse D,  Johnson T;
XX
DR   WPI; 1995-224291/29.
DR   P-PSDB; AAR75458.
XX
PT   Generating new antibodies specific for immunorecessive epitopes -
PT   by selection from variegated V gene library cloned from
PT   immuno:tolerance derived antibody repertoire, useful in diagnosis,
PT   purifcn. and therapy, e.g. of cancer
XX
PS   Disclosure; Page 77-78; 109pp; English.
XX
CC   The coding sequence of the heavy chain variable region from the mouse
CC   antibody F4-7. This sequence was isolated from a variegated display
CC   library (VDL) of variable regions derived from a repertoire of
CC   antibodies from an immunotolerised animal.  The VDL is generated by PCR
CC   amplifying the variable regions from the antibody coding sequences using
CC   the primers AAQ74153-74. The variable regions, esp the complementarity
CC   determining regions (CDR; see AAR75462-93 for examples of CDRs) from the
CC   immunotolerant animals' antibodies are used to construct an antibody
CC   against a immunorecessive antigen e.g. a cell surface marker on a foetal,
CC   cancer or stem cell, which can differentiate between variant or related
CC   forms of the antigen.  The antibodies generated can be used in the
CC   diagnosis, e.g. detection of the immunorecessive antigen, or in therapy
CC   e.g. of cancer, Alzheimer's disease or familial hypercholesterolaemia.
CC   The method of production of the antibody allows rapid and sensitive
CC   isolation of antibodies that would be difficult to isolate by standard
CC   methods.  The antibodies produced have greater binding affinity than
CC   those produced by combinatorial/hybridoma methods.
XX
SQ   Sequence 735 BP; 172 A; 213 C; 188 G; 162 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 16;  Length 735;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db     465 tgctggatctgctg 478


RESULT  38
AAZ00347/c
ID   AAZ00347 standard; DNA; 753 BP.
XX
AC   AAZ00347;
XX
DT   26-OCT-1999  (first entry)
XX
DE   Nucleotide sequence of human hsFATP3.
XX
KW   Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW   fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9936537-A2.
XX
PD   22-JUL-1999.
XX
PF   14-JAN-1999;   99WO-US00182.
XX
PR   14-JAN-1999;   99US-0232201.
PR   15-JAN-1998;   98US-0071374.
PR   20-JUL-1998;   98US-0093491.
PR   04-DEC-1998;   98US-0110941.
PR   14-JAN-1999;   99US-0232195.
PR   14-JAN-1999;   99US-0232197.
PR   14-JAN-1999;   99US-0232200.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Gimeno RE,  Hirsch DJ,  Lodish HF,  Stahl A,  Tartaglia LA;
XX
DR   WPI; 1999-444398/37.
DR   P-PSDB; AAY14937.
XX
PT   Fatty acid transport proteins and related polynucleotides, useful
PT   for treating obesity, diabetes and heart disease
XX
PS   Example 1; Fig 16; 255pp; English.
XX
CC   The invention provides a family of fatty acid transport proteins (FATPs)
CC   that mediate transport of long chain fatty acids (LCFAs) across cell
CC   membranes into cells. Human and murine FATP proteins and nucleic acids
CC   encoding the proteins are provided. The FATP proteins can be produced
CC   by standard recombinant methodology. Fatty acid uptake by cells can be
CC   modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC   In particular, antisense oligonucleotides can be used to modulate FATP
CC   biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
```

CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
SQ Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 753;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tctgctgtttgaag 23
||||||||||||||
Db 436 TCTGCTGTTTGAAG 423

RESULT 39
AAF89005/c
ID AAF89005 standard; DNA; 753 BP.
XX
AC AAF89005;
XX
DT 06-JUL-2001 (first entry)
XX
DE Human FATP3 coding sequence SEQ ID NO: 14.
XX
KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
KW weight control; tuberculosis; TB; anti-fungal; ds.
XX
OS Homo sapiens.
XX
PN WO200121795-A2.
XX
PD 29-MAR-2001.
XX
PF 21-SEP-2000; 2000WO-US25891.
XX
PR 23-SEP-1999; 99US-0405504.
PR 23-SEP-1999; 99US-0405505.
PR 16-DEC-1999; 99US-0465280.
PR 17-FEB-2000; 2000US-0506252.
PR 06-JUL-2000; 2000US-0611197.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;
XX
DR WPI; 2001-354783/37.
DR P-PSDB; AAB83227.
XX
PT New fatty acid transport proteins (FATPs) useful for the manufacture of
PT medicament for treating obesity, diabetes and heart disease -
XX
PS Disclosure; Fig 16; 287pp; English.
XX
CC The present invention provides the protein and coding sequences of fatty
CC acid transport proteins (FATPs) from a number of species, including
CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
CC from the mouse, FATPa and b from C. elegans, and FATP from Aspergillus
CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
CC tuberculosis can be used to identify inhibitors which can then be used to
CC treat TB. That from M. grisea (also known as rice blast fungus) can be
CC used to develop anti-fungal agents capable of preventing infection of
CC rice. Those from the human can be used to develop treatments for
CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
CC present sequence is one of the sequences described in the exemplification
CC of the invention.
XX
SQ Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 753;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tctgctgtttgaag 23
||||||||||||||
Db 436 TCTGCTGTTTGAAG 423

RESULT 40
AAF15012/c
ID AAF15012 standard; cDNA; 769 BP.
XX
AC AAF15012;
XX
DT 13-MAR-2001 (first entry)
XX
DE Trichoderma reesei EST SEQ ID NO:7535.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Trichoderma reesei.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 89; Page 3043-3044; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the

```
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC   Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC   niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC   AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC   all specifically claimed in the present invention.
XX
SQ   Sequence 769 BP; 203 A; 234 C; 203 G; 114 T; 15 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 769;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db    284 TGCTGGATCTGCTG 271


RESULT  41
AAZ06233
ID   AAZ06233 standard; DNA; 891 BP.
XX
AC   AAZ06233;
XX
DT   30-SEP-1999  (first entry)
XX
DE   Human secreted protein gene No. 15.
XX
KW   Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW   diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW   developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW   immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW   inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW   cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW   osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW   endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS   Homo sapiens.
XX
PN   WO9935158-A1.
XX
PD   15-JUL-1999.
XX
PF   06-JAN-1999;   99WO-US00108.
XX
PR   07-JAN-1998;   98US-0070704.
PR   07-JAN-1998;   98US-0070657.
PR   07-JAN-1998;   98US-0070658.
PR   07-JAN-1998;   98US-0070692.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Brewer LA,  Duan RD,  Ebner R,  Lafleur DW,  Ni J;
PI   Olsen HS,  Rosen CA,  Ruben SM,  Soppet DR;
XX
DR   WPI; 1999-444190/37.
DR   P-PSDB; AAY38400, AAY38453.
XX
PT   New isolated human genes and the secreted polypeptides they encode
XX
PS   Claim 1; Page 158; 227pp; English.
XX
CC   This sequence represents a nucleic acid molecule which encodes a
CC   secreted human protein. The gene number is given in the descriptor line.
CC   The gene can be used to generate fusion proteins by linking to the gene
CC   to a human immunoglobulin Fc portion (e.g. AAZ06210) for increasing the
CC   stability of the fused protein as compared to the human protein only.
CC   The invention relates to 36 novel genes and their fragments (nucleic
CC   acid sequences: AAZ06219-Z06263; amino acid sequences AAY38386-Y38498)
CC   which are useful for preventing, treating or ameliorating  medical
CC   conditions e.g. by protein or gene therapy.  Also, pathological
CC   conditions can be diagnosed by determining the amount of the new
CC   polypeptides in a sample or by determining the presence of mutations in
CC   the new polynucleotides. Specific uses are described for each of the 36
CC   polynucleotides, based on which tissues they are most highly expressed in
CC   (see AAZ06219 for described uses).
XX
SQ   Sequence 891 BP; 207 A; 214 C; 220 G; 250 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 891;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctggatctgct 15
          ||||||||||||||
Db    727 ttgctggatctgct 740


RESULT  42
AAX82834/c
ID   AAX82834 standard; DNA; 1044 BP.
XX
AC   AAX82834;
XX
DT   30-JUN-2000  (first entry)
XX
DE   L-asparaginase DNA.
XX
KW   L-asparaginase; fermentation; ss.
XX
OS   Unidentified.
XX
PN   CN1237633-A.
XX
PD   08-DEC-1999.
XX
PF   01-JUN-1998;   98CN-0102046.
XX
PR   01-JUN-1998;   98CN-0102046.
XX
PA   (MICR-) INST MICROBIOLOGY CHINESE ACAD SCI.
XX
PI   Qian S,  Wang Y,  Meng G;
XX
DR   WPI; 2000-351194/31.
DR   P-PSDB; AAW90867.
XX
PT   New structured L-asparaginase bacterial host cell and its production
PT   culture -
XX
PS   Disclosure; Fig 2; 12pp; Chinese.
XX
CC   This invention describes a novel L-asparaginase-producing recombinant
CC   bacteria and its fermentation culture method. The new cell is produced
CC   using standard recombinant techniques. The enzyme gene expressed in
CC   the cell is different from the reported enzyme gene in sequence, and
CC   its individual amino acid composition is also different. The culture
CC   medium adopted for culturing the new bacterial cells is cheap and easily
CC   available, and the fermentation process used to grow the cells and
CC   express the enzyme is simple and easy to implement. This sequence
CC   encodes an L-asparaginase which is described in the method of the
CC   invention.
XX
SQ   Sequence 1044 BP; 274 A; 281 C; 272 G; 217 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 1044;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   1028 TGCTGGATCTGCTG 1015
```

```
RESULT  43
AAZ09784
ID   AAZ09784 standard; DNA; 1155 BP.
XX
AC   AAZ09784;
XX
DT   23-NOV-1999  (first entry)
XX
DE   E. coli S-adenosyl methionine synthase SAM DNA.
XX
KW   SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bioS1; bioS2;
KW   bioS3; cofactor; decarboxylation; Vitamin H; ss.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1155
FT                   /*tag= a
FT                   /product= "S-adenosyl methionine synthase"
XX
PN   DE19806872-A1.
XX
PD   26-AUG-1999.
XX
PF   19-FEB-1998;   98DE-1006872.
XX
PR   19-FEB-1998;   98DE-1006872.
XX
PA   (BADI ) BASF AG.
XX
PI   Schroeder H;
XX
DR   WPI; 1999-480095/41.
DR   P-PSDB; AAY33263.
XX
PT   Production of biotin by expressing S-adenosyl-methionine synthase and
PT   second biotin synthesis gene in host cells -
XX
PS   Claim 1; Page 8-10; 48pp; German.
XX
CC   This invention describes a novel method for the preparation of biotin
CC   (I) which comprises expressing, in a prokaryotic or eukaryotic host
CC   capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
CC   sequence (1), and (b) at least one of the other biotin biosynthesis
CC   genes bioS1, 2 or 3. (I) is a cofactor for enzyme-catalyzed
CC   (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC   most animals and some microorganisms. Expression of biotin plus bioS1,
CC   bioS2 or bioS3 significantly increases productivity of biotin
CC   biosynthesis, particularly by at least 3 times. This sequence encodes
CC   the Escherichia coli SAM protein which is used in the method of the
CC   invention.
XX
SQ   Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 1155;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db    1031 tgctggatctgctg 1044


RESULT  44
AAA99433
ID   AAA99433 standard; DNA; 1155 BP.
XX
AC   AAA99433;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Met K oligonucleotide SEQ ID 17 used in preparation of L-methionine.
XX
KW   L-methionine production; Met; Escherichia coli; ds.
XX
OS   Escherichia coli.
XX
PN   JP2000139471-A.
XX
PD   23-MAY-2000.
XX
PF   17-NOV-1998;   98JP-0326717.
XX
PR   17-NOV-1998;   98JP-0326717.
XX
PA   (AJIN ) AJINOMOTO KK.
XX
DR   WPI; 2001-018703/03.
DR   P-PSDB; AAB26814.
XX
PT   Fermentative process for preparation of L-methionine (Met), comprises
PT   using a modified Met producing microorganism, particularly an
PT   Escherichia species -
XX
PS   Example 2; Page 17-18; 23pp; Japanese.
XX
CC   The invention relates to a process for preparing L-methionine (Met)
CC   using a modified Met producing organism. The modified microorganism has
CC   its Met biosynthetic system repressor deleted, and has enhanced
CC   homoserine transsuccunylase activity. The organism optionally has an
CC   attenuated intracellular S-adenosyl methionine synthetase (SAM). The
CC   process is used for the preparation of L-methionine. Sequences AAA99416
CC   and AAA99433 encode proteins represented in AAB26816 and AAB26814
CC   respectively, which are involved in the Escherichia coli Met biosynthetic
CC   pathway. AAA99417-A99441 (excluding AAA99433) represent primers, all
CC   sequences are used in the production of the modified Met producing
CC   organism of the invention.
XX
SQ   Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 1155;
  Best Local Similarity   100.0%;  Pred. No. 1.1e+02;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db    1031 tgctggatctgctg 1044


RESULT  45
AAZ00358/c
ID   AAZ00358 standard; DNA; 1173 BP.
XX
AC   AAZ00358;
XX
DT   26-OCT-1999  (first entry)
XX
DE   Nucleotide sequence of human hsFATP3.
XX
KW   Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW   fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9936537-A2.
XX
PD   22-JUL-1999.
XX
PF   14-JAN-1999;   99WO-US00182.
XX
PR   14-JAN-1999;   99US-0232201.
PR   15-JAN-1998;   98US-0071374.
PR   20-JUL-1998;   98US-0093491.
```

PR 04-DEC-1998; 98US-0110941.
PR 14-JAN-1999; 99US-0232195.
PR 14-JAN-1999; 99US-0232197.
PR 14-JAN-1999; 99US-0232200.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
DR WPI; 1999-444398/37.
DR P-PSDB; AAY14948.
XX
PT Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
PS Claim 31; Fig 48; 255pp; English.
XX
CC The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
SQ Sequence 1173 BP; 247 A; 327 C; 351 G; 248 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 1173;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 tctgctgtttgaag 23
||||||||||||||
Db 853 TCTGCTGTTTGAAG 840

Search completed: December 26, 2001, 14:16:34
Job time: 6986 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:52:31 ; Search time 143.1 Seconds
(without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196F-3
Perfect score: 25
Sequence: 1 gttgctggatctgctgtttgaagcg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 9

Total number of hits satisfying chosen parameters: 14184

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
2		25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
3		25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
4		25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
5		16	64.0	5000	3	US-09-104-070-1	Sequence 1, Appli
6	c	15	60.0	708	4	US-08-998-416-997	Sequence 997, App
7	c	14	56.0	384	2	US-08-820-170A-8	Sequence 8, Appli
8	c	14	56.0	384	3	US-09-055-699-8	Sequence 8, Appli
9	c	14	56.0	384	4	US-09-273-565-8	Sequence 8, Appli
10		14	56.0	735	5	PCT-US94-14106-54	Sequence 54, Appl
11	c	14	56.0	753	4	US-09-232-191-14	Sequence 14, Appl
12	c	14	56.0	753	4	US-09-232-200-14	Sequence 14, Appl
13	c	14	56.0	753	4	US-09-232-197-14	Sequence 14, Appl
14	c	14	56.0	1173	4	US-09-232-200-50	Sequence 50, Appl
15	c	14	56.0	1173	4	US-09-232-197-50	Sequence 50, Appl
16	c	14	56.0	1493	2	US-08-820-170A-9	Sequence 9, Appli
17	c	14	56.0	1493	3	US-09-055-699-9	Sequence 9, Appli
18	c	14	56.0	1493	4	US-09-273-565-9	Sequence 9, Appli
19	c	14	56.0	1848	1	US-08-447-422-15	Sequence 15, Appl
20	c	14	56.0	1998	4	US-09-232-200-68	Sequence 68, Appl
21	c	14	56.0	1998	4	US-09-232-197-68	Sequence 68, Appl
22	c	14	56.0	2087	4	US-09-232-191-6	Sequence 6, Appli
23	c	14	56.0	2087	4	US-09-232-200-6	Sequence 6, Appli
24	c	14	56.0	2087	4	US-09-232-197-6	Sequence 6, Appli
25	c	14	56.0	2166	4	US-09-232-200-101	Sequence 101, App
26	c	14	56.0	2166	4	US-09-232-197-101	Sequence 101, App
27	c	14	56.0	3243	2	US-08-239-276-6	Sequence 6, Appli
28	c	14	56.0	3243	2	US-08-468-579B-6	Sequence 6, Appli
29	c	14	56.0	3243	3	US-08-468-577B-6	Sequence 6, Appli
30	c	14	56.0	3282	1	US-08-072-574-11	Sequence 11, Appl
31	c	14	56.0	3282	1	US-08-486-270-11	Sequence 11, Appl
32	c	14	56.0	3282	3	US-08-367-264-11	Sequence 11, Appl
33	c	14	56.0	4085	1	US-08-072-574-7	Sequence 7, Appli
34	c	14	56.0	4085	1	US-08-486-270-7	Sequence 7, Appli
35	c	14	56.0	4085	3	US-08-367-264-7	Sequence 7, Appli
36	c	14	56.0	4181	1	US-08-072-574-9	Sequence 9, Appli
37	c	14	56.0	4181	1	US-08-486-270-9	Sequence 9, Appli
38	c	14	56.0	4181	3	US-08-367-264-9	Sequence 9, Appli
39	c	14	56.0	4207	4	US-08-660-148-1	Sequence 1, Appli
40	c	14	56.0	4207	4	US-08-660-148-3	Sequence 3, Appli
41	c	14	56.0	4303	4	US-08-660-148-4	Sequence 4, Appli
42	c	14	56.0	4303	4	US-08-660-148-6	Sequence 6, Appli
43	c	14	56.0	4566	2	US-08-465-976A-1	Sequence 1, Appli
44	c	14	56.0	4566	2	US-08-982-412-1	Sequence 1, Appli
45		14	56.0	5319	1	US-08-169-927-1	Sequence 1, Appli
46	c	14	56.0	7326	1	US-08-194-468-1	Sequence 1, Appli
47	c	14	56.0	7344	3	US-08-961-739-1	Sequence 1, Appli
48		14	56.0	12284	2	US-08-876-991-1	Sequence 1, Appli
49		14	56.0	12284	2	US-09-059-853-1	Sequence 1, Appli
50	c	14	56.0	28804	2	US-08-592-874-1	Sequence 1, Appli
51	c	14	56.0	28804	3	US-09-096-942-2	Sequence 2, Appli
52	c	14	56.0	28804	3	US-09-096-867-2	Sequence 2, Appli
53		13	52.0	749	4	US-08-998-416-727	Sequence 727, App
54		13	52.0	1161	1	US-08-153-848-31	Sequence 31, Appl
55		13	52.0	1161	5	PCT-US93-11153-31	Sequence 31, Appl
56	c	13	52.0	1242	1	US-08-252-966B-13	Sequence 13, Appl
57	c	13	52.0	1729	3	US-09-188-930-68	Sequence 68, Appl
58		13	52.0	1835	1	US-08-361-611-3	Sequence 3, Appli
59		13	52.0	1835	1	US-08-565-655-3	Sequence 3, Appli
60		13	52.0	1835	2	US-08-946-967-3	Sequence 3, Appli
61		13	52.0	1862	3	US-08-999-723-1	Sequence 1, Appli
62		13	52.0	1862	4	US-09-434-427-1	Sequence 1, Appli
63		13	52.0	2143	4	US-09-064-693A-18	Sequence 18, Appl
64		13	52.0	2254	1	US-08-153-848-27	Sequence 27, Appl
65		13	52.0	2254	3	US-09-299-843A-27	Sequence 27, Appl
66		13	52.0	2254	5	PCT-US93-11153-27	Sequence 27, Appl
67		13	52.0	2492	4	US-09-381-779-1	Sequence 1, Appli
68	c	13	52.0	2674	2	US-08-926-724-2	Sequence 2, Appli
69		13	52.0	3003	2	US-08-873-479-41	Sequence 41, Appl
70		13	52.0	3119	3	US-09-299-843A-31	Sequence 31, Appl
71	c	13	52.0	3206	1	US-08-373-134D-1	Sequence 1, Appli
72	c	13	52.0	3206	2	US-09-114-637-1	Sequence 1, Appli
73		13	52.0	3435	1	US-08-366-577-1	Sequence 1, Appli
74		13	52.0	3435	5	PCT-US96-00005-1	Sequence 1, Appli
75		13	52.0	3582	2	US-08-400-159-9	Sequence 9, Appli
76		13	52.0	3582	3	US-08-611-729A-9	Sequence 9, Appli
77	c	13	52.0	3603	1	US-08-188-582-15	Sequence 15, Appl
78	c	13	52.0	3603	1	US-08-646-715-15	Sequence 15, Appl
79	c	13	52.0	3731	4	US-08-811-583-1	Sequence 1, Appli
80	c	13	52.0	4800	3	US-09-106-638-1	Sequence 1, Appli
81		13	52.0	4856	4	US-09-045-360-1	Sequence 1, Appli
82	c	13	52.0	4954	1	US-08-920-812-18	Sequence 18, Appl
83	c	13	52.0	4954	1	US-08-920-827-18	Sequence 18, Appl
84	c	13	52.0	4954	1	US-08-921-177-18	Sequence 18, Appl
85	c	13	52.0	4954	1	US-08-362-577C-18	Sequence 18, Appl
86	c	13	52.0	4954	2	US-08-920-828-18	Sequence 18, Appl
87		13	52.0	6641	4	US-09-064-693A-25	Sequence 25, Appl
88		13	52.0	7032	2	US-08-149-097D-24	Sequence 24, Appl
89		13	52.0	7032	3	US-08-949-386-24	Sequence 24, Appl
90		13	52.0	7032	3	US-08-450-562-24	Sequence 24, Appl
91		13	52.0	7089	3	US-08-949-386-25	Sequence 25, Appl
92		13	52.0	7089	3	US-08-450-562-25	Sequence 25, Appl
93	c	13	52.0	8878	1	US-08-759-444-2	Sequence 2, Appli
94	c	13	52.0	9880	3	US-08-680-897-1	Sequence 1, Appli
95	c	13	52.0	24979	2	US-08-147-777-3	Sequence 3, Appli
96	c	13	52.0	24979	3	US-08-452-872-3	Sequence 3, Appli
97	c	13	52.0	24979	5	PCT-US93-03985-3	Sequence 3, Appli
98	c	13	52.0	46899	1	US-08-471-119A-1	Sequence 1, Appli
99	c	13	52.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
100	c	13	52.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db      57 GTTGCTGGATCTGCTGTTTGAAGCG 81


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db      57 GTTGCTGGATCTGCTGTTTGAAGCG 81


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db     173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgtttgaagcg 25
           |||||||||||||||||||||||||
Db     173 GTTGCTGGATCTGCTGTTTGAAGCG 197


RESULT   5
US-09-104-070-1
; Sequence 1, Application US/09104070
; Patent No. 6153741
; GENERAL INFORMATION:
;  APPLICANT: Richards, Eric J.
;  APPLICANT: Jeddeloh, Jeffrey A.
;  TITLE OF INVENTION: DNA Methylation Gene from Plants
;  FILE REFERENCE: Wash U CI-0014 US
;  CURRENT APPLICATION NUMBER: US/09/104,070
;  CURRENT FILING DATE: 1998-06-24
;  EARLIER APPLICATION NUMBER: US 60/083,612
;  EARLIER FILING DATE: 1998-04-30
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 5000
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (782)...(1252)
;   NAME/KEY: CDS
;   LOCATION: (1354)...(1440)
;   NAME/KEY: CDS
;   LOCATION: (1549)...(1895)
;   NAME/KEY: CDS
;   LOCATION: (1976)...(2165)
;   NAME/KEY: CDS
;   LOCATION: (2251)...(2437)
;   NAME/KEY: CDS
;   LOCATION: (2559)...(2629)
;   NAME/KEY: CDS
;   LOCATION: (2703)...(2892)
;   NAME/KEY: CDS
;   LOCATION: (2975)...(3070)
;   NAME/KEY: CDS
;   LOCATION: (3148)...(3242)
;   NAME/KEY: CDS
;   LOCATION: (3317)...(3436)
;   NAME/KEY: CDS
;   LOCATION: (3540)...(3659)
;   NAME/KEY: CDS
;   LOCATION: (3745)...(3843)
;   NAME/KEY: CDS
;   LOCATION: (3934)...(4038)
;   NAME/KEY: CDS
;   LOCATION: (4130)...(4354)
;   NAME/KEY: CDS
;   LOCATION: (4826)...(4755)
;   OTHER INFORMATION: t-RNA-glu coding region
;   FEATURE:
;   NAME/KEY: mutation
;   LOCATION: (785)...(786)
;   OTHER INFORMATION: som8 rearrangement, deletion of G at 785 and
;   OTHER INFORMATION: insertion of 83 bp between 785 and 786
;   FEATURE:
;   NAME/KEY: mutation
;   LOCATION: (3243)...(3243)
;   OTHER INFORMATION: ddm1-2 base pair substitution of G to A
US-09-104-070-1


  Query Match             64.0%;  Score 16;  DB 3;  Length 5000;
  Best Local Similarity   100.0%;  Pred. No. 2;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctgtt 18
           ||||||||||||||||
Db    1132 tgctggatctgctgtt 1147


RESULT   6
US-08-998-416-997/c
; Sequence 997, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  997:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  708 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1612RP
US-08-998-416-997


  Query Match             60.0%;  Score 15;  DB 4;  Length 708;
  Best Local Similarity   100.0%;  Pred. No. 6.8;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctgt 17
           |||||||||||||||
Db     470 TGCTGGATCTGCTGT 456


RESULT   7
US-08-820-170A-8/c
; Sequence 8, Application US/08820170A
; Patent No. 5831058
;  GENERAL INFORMATION:
;    APPLICANT:  Tsutomu, FUJIWARA
;    APPLICANT:  Takeshi, WATANABE
;    APPLICANT:  Masato, HORIE
;    APPLICANT:  Toyomasa, KATAGIRI
;    TITLE OF INVENTION:  HUMAN GENE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/820,170A
;      FILING DATE:
;      CLASSIFICATION:  536
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  384 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA(genomic)
US-08-820-170A-8


  Query Match             56.0%;  Score 14;  DB 2;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tggatctgctgttt 19
           ||||||||||||||
Db     350 TGGATCTGCTGTTT 337


RESULT   8
US-09-055-699-8/c
; Sequence 8, Application US/09055699
; Patent No. 6005088
;  GENERAL INFORMATION:
;    APPLICANT:  Tsutomu, FUJIWARA
;    APPLICANT:  Takeshi, WATANABE
;    APPLICANT:  Masato, HORIE
;    APPLICANT:  Toyomasa, KATAGIRI
;    TITLE OF INVENTION:  HUMAN GENE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/055,699
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/820,170
;      FILING DATE:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  384 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA(genomic)
US-09-055-699-8


  Query Match             56.0%;  Score 14;  DB 3;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tggatctgctgttt 19
           ||||||||||||||
Db     350 TGGATCTGCTGTTT 337


RESULT   9
US-09-273-565-8/c
; Sequence 8, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
;  APPLICANT: FUJIWARA, TSUTOMU
;  APPLICANT: WATANABE, TAKESHI
;  APPLICANT: HORIE, MASATO
;  TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
;  TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
;  FILE REFERENCE: Q-53599
;  CURRENT APPLICATION NUMBER: US/09/273,565A
;  CURRENT FILING DATE: 1999-03-22
;  EARLIER APPLICATION NUMBER: 09/055,699
;  EARLIER FILING DATE: 1998-04-07
;  EARLIER APPLICATION NUMBER: 08/820,170
;  EARLIER FILING DATE: 1997-03-19
;  EARLIER APPLICATION NUMBER: JP 63410/1996
;  EARLIER FILING DATE: 1996-03-19
;  EARLIER APPLICATION NUMBER: JP 69163/1997
;  EARLIER FILING DATE: 1997-03-05
;  NUMBER OF SEQ ID NOS: 95
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 384
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-273-565-8


  Query Match             56.0%;  Score 14;  DB 4;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tggatctgctgttt 19
           ||||||||||||||
Db     350 TGGATCTGCTGTTT 337


RESULT  10
PCT-US94-14106-54
; Sequence 54, Application PC/TUS9414106
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Process for Generating Specific Antibodies
;    NUMBER OF SEQUENCES:  61
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII (text)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/14106
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  54:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  735 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  67..735
PCT-US94-14106-54


  Query Match             56.0%;  Score 14;  DB 5;  Length 735;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db     465 TGCTGGATCTGCTG 478


RESULT  11
US-09-232-191-14/c
; Sequence 14, Application US/09232191
; Patent No. 6284487
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  TITLE OF INVENTION: Fatty Acid Transport Proteins
;  FILE REFERENCE: WHI97-21p3ME
;  CURRENT APPLICATION NUMBER: US/09/232,191
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-232-191-14


  Query Match             56.0%;  Score 14;  DB 4;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db     436 TCTGCTGTTTGAAG 423


RESULT  12
US-09-232-200-14/c
; Sequence 14, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
```

```
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-232-200-14


  Query Match             56.0%;  Score 14;  DB 4;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaaag 23
           |||||||||||||||
Db     436 TCTGCTGTTTGAAG 423


RESULT  13
US-09-232-197-14/c
; Sequence 14, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-232-197-14


  Query Match             56.0%;  Score 14;  DB 4;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaaag 23
           |||||||||||||||
Db     436 TCTGCTGTTTGAAG 423


RESULT  14
US-09-232-200-50/c
; Sequence 50, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 50
;   LENGTH: 1173
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-232-200-50


  Query Match             56.0%;  Score 14;  DB 4;  Length 1173;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaaag 23
           |||||||||||||||
Db     853 TCTGCTGTTTGAAG 840


RESULT  15
US-09-232-197-50/c
; Sequence 50, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 50
;   LENGTH: 1173
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-232-197-50


  Query Match             56.0%;  Score 14;  DB 4;  Length 1173;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaaag 23
           |||||||||||||||
Db     853 TCTGCTGTTTGAAG 840


RESULT  16
US-08-820-170A-9/c
; Sequence 9, Application US/08820170A
; Patent No. 5831058
;  GENERAL INFORMATION:
;    APPLICANT:  Tsutomu, FUJIWARA
;    APPLICANT:  Takeshi, WATANABE
;    APPLICANT:  Masato, HORIE
;    APPLICANT:  Toyomasa, KATAGIRI
;    TITLE OF INVENTION:  HUMAN GENE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
```

```
;       STREET:  2100 Pennsylvania Avenue, N.W.
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  United States
;       ZIP:  20037-3202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/820,170A
;       FILING DATE:
;       CLASSIFICATION:  536
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 293-7060
;       TELEFAX:  (202) 293-7860
;       TELEX:  6491103
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1493 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA(genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     IMMEDIATE SOURCE:
;       LIBRARY:  Human fetal brain cDNA library
;       CLONE:  GEN-025F07
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  95..478
US-08-820-170A-9


  Query Match             56.0%;  Score 14;  DB 2;  Length 1493;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tggatctgctgttt 19
           ||||||||||||||
Db     444 TGGATCTGCTGTTT 431


RESULT  17
US-09-055-699-9/c
; Sequence 9, Application US/09055699
; Patent No. 6005088
;  GENERAL INFORMATION:
;     APPLICANT:  Tsutomu, FUJIWARA
;     APPLICANT:  Takeshi, WATANABE
;     APPLICANT:  Masato, HORIE
;     APPLICANT:  Toyomasa, KATAGIRI
;     TITLE OF INVENTION:  HUMAN GENE
;     NUMBER OF SEQUENCES:  42
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;       STREET:  2100 Pennsylvania Avenue, N.W.
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  United States
;       ZIP:  20037-3202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/055,699
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/820,170
;       FILING DATE:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 293-7060
;       TELEFAX:  (202) 293-7860
;       TELEX:  6491103
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1493 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA(genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     IMMEDIATE SOURCE:
;       LIBRARY:  Human fetal brain cDNA library
;       CLONE:  GEN-025F07
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  95..478
US-09-055-699-9


  Query Match             56.0%;  Score 14;  DB 3;  Length 1493;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tggatctgctgttt 19
           ||||||||||||||
Db     444 TGGATCTGCTGTTT 431


RESULT  18
US-09-273-565-9/c
; Sequence 9, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
;  APPLICANT: FUJIWARA, TSUTOMU
;  APPLICANT: WATANABE, TAKESHI
;  APPLICANT: HORIE, MASATO
;  TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
;  TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
;  FILE REFERENCE: Q-53599
;  CURRENT APPLICATION NUMBER: US/09/273,565A
;  CURRENT FILING DATE: 1999-03-22
;  EARLIER APPLICATION NUMBER: 09/055,699
;  EARLIER FILING DATE: 1998-04-07
;  EARLIER APPLICATION NUMBER: 08/820,170
;  EARLIER FILING DATE: 1997-03-19
;  EARLIER APPLICATION NUMBER: JP 63410/1996
;  EARLIER FILING DATE: 1996-03-19
;  EARLIER APPLICATION NUMBER: JP 69163/1997
;  EARLIER FILING DATE: 1997-03-05
;  NUMBER OF SEQ ID NOS: 95
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 1493
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (95)..(478)
US-09-273-565-9


  Query Match             56.0%;  Score 14;  DB 4;  Length 1493;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tggatctgctgttt 19
           ||||||||||||||
Db     444 TGGATCTGCTGTTT 431
```

```
RESULT  19
US-08-447-422-15/c
; Sequence 15, Application US/08447422
; Patent No. 5686579
;  GENERAL INFORMATION:
;    APPLICANT:  SHAMI, Ezekiel Y.
;    APPLICANT:  ROTHSTEIN, Aser
;    APPLICANT:  RAMJEESINGH, Mohabir
;    TITLE OF INVENTION:  Use of Antibody/Antigen Interactions To
;    TITLE OF INVENTION:  Protect or Modulate Biological Activity
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/447,422
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/081,410
;      FILING DATE:  22-JUN-1993
;      APPLICATION NUMBER:  US 07/938,505
;      FILING DATE:  31-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/205,748
;      FILING DATE:  21-JUN-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  17923/102 HYLI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1848 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-447-422-15


  Query Match             56.0%;  Score 14;  DB 1;  Length 1848;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   1829 TGCTGGATCTGCTG 1816


RESULT  20
US-09-232-200-68/c
; Sequence 68, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 68
;   LENGTH: 1998
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-200-68


  Query Match             56.0%;  Score 14;  DB 4;  Length 1998;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db   1690 TCTGCTGTTTGAAG 1677


RESULT  21
US-09-232-197-68/c
; Sequence 68, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 68
;   LENGTH: 1998
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-197-68


  Query Match             56.0%;  Score 14;  DB 4;  Length 1998;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db   1690 TCTGCTGTTTGAAG 1677


RESULT  22
US-09-232-191-6/c
; Sequence 6, Application US/09232191
; Patent No. 6284487
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
```

```
;  TITLE OF INVENTION: Fatty Acid Transport Proteins
;  FILE REFERENCE: WHI97-21p3ME
;  CURRENT APPLICATION NUMBER: US/09/232,191
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 2087
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-191-6


  Query Match           56.0%;  Score 14;  DB 4;  Length 2087;
  Best Local Similarity  100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db    1779 TCTGCTGTTTGAAG 1766


RESULT  23
US-09-232-200-6/c
; Sequence 6, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 2087
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-200-6


  Query Match           56.0%;  Score 14;  DB 4;  Length 2087;
  Best Local Similarity  100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db    1779 TCTGCTGTTTGAAG 1766


RESULT  24
US-09-232-197-6/c
; Sequence 6, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 2087
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-197-6


  Query Match           56.0%;  Score 14;  DB 4;  Length 2087;
  Best Local Similarity  100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db    1779 TCTGCTGTTTGAAG 1766


RESULT  25
US-09-232-200-101/c
; Sequence 101, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 101
;   LENGTH: 2166
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (19)...(2124)
US-09-232-200-101


  Query Match           56.0%;  Score 14;  DB 4;  Length 2166;
  Best Local Similarity  100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db    1987 TCTGCTGTTTGAAG 1974


RESULT  26
```

```
US-09-232-197-101/c
; Sequence 101, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 101
;   LENGTH: 2166
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (19)...(2124)
US-09-232-197-101


  Query Match             56.0%;  Score 14;  DB 4;  Length 2166;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaag 23
           ||||||||||||||
Db    1987 TCTGCTGTTTGAAG 1974


RESULT  27
US-08-239-276-6/c
; Sequence 6, Application US/08239276
; Patent No. 5840836
;  GENERAL INFORMATION:
;    APPLICANT:  Rabin, Daniel
;    TITLE OF INVENTION:  PANCREATIC ISLET CELL ANTIGENS
;    TITLE OF INVENTION:  OBTAINED BY MOLECULAR CLONING
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  System 7.5
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/239,276
;      FILING DATE:  05-MAY-1994
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/872,646
;      FILING DATE:  08-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/715,181
;      FILING DATE:  14-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/441,703
;      FILING DATE:  04-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/312,543
;      FILING DATE:  17-FEB-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  MDI 251.4-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3243 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-239-276-6


  Query Match             56.0%;  Score 14;  DB 2;  Length 3243;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctqctg 16
             ||||||||||||||
Db     413 TGCTGGATCTGCTG 400


RESULT  28
US-08-468-579B-6/c
; Sequence 6, Application US/08468579B
; Patent No. 5981700
;  GENERAL INFORMATION:
;    APPLICANT:  Rabin, Daniel
;    TITLE OF INVENTION:  PANCREATIC ISLET CELL ANTIGENS
;    TITLE OF INVENTION:  OBTAINED BY MOLECULAR CLONING
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  System 7.5
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,579B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,276
;      FILING DATE:  05-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/872,646
;      FILING DATE:  08-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/715,181
;      FILING DATE:  14-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/441,703
;      FILING DATE:  04-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/312,543
;      FILING DATE:  17-FEB-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  MDI 251.5-KGB
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3243 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-468-579B-6


  Query Match            56.0%;  Score 14;  DB 2;  Length 3243;
  Best Local Similarity  100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db     413 TGCTGGATCTGCTG 400


RESULT  29
US-08-468-577B-6/c
; Sequence 6, Application US/08468577B
; Patent No. 6001804
;  GENERAL INFORMATION:
;    APPLICANT:  Rabin, Daniel
;    TITLE OF INVENTION:  PANCREATIC ISLET CELL ANTIGENS
;    TITLE OF INVENTION:  OBTAINED BY MOLECULAR CLONING
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sprung Kramer Schaefer & Briscoe
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4 Mb storage
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  System 7.5
;      SOFTWARE:  WordPerfect
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,577B
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/239,276
;      FILING DATE:  05-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/872,646
;      FILING DATE:  08-JUN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/715,181
;      FILING DATE:  14-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/441,703
;      FILING DATE:  04-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/312,543
;      FILING DATE:  17-FEB-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  MDI 251.8-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3243 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-468-577B-6


  Query Match            56.0%;  Score 14;  DB 3;  Length 3243;
  Best Local Similarity  100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db     413 TGCTGGATCTGCTG 400


RESULT  30
US-08-072-574-11/c
; Sequence 11, Application US/08072574
; Patent No. 5521297
;  GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen
;    APPLICANT:  Pontsler, Aaron
;    TITLE OF INVENTION:  HUMAN MATABOTROPIC GLUTAMATE RECEPTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  00719
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,574
;      FILING DATE:  19930604
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9383
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  370..3003
;      OTHER INFORMATION:  /product= "HUMAN MGLUR5C"
;      OTHER INFORMATION:  /note= "Variant of MGLUR5A with truncated 3' end."
US-08-072-574-11


  Query Match            56.0%;  Score 14;  DB 1;  Length 3282;
  Best Local Similarity  100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db    2950 TGCTGGATCTGCTG 2937


RESULT  31
```

```
US-08-486-270-11/c
; Sequence 11, Application US/08486270
; Patent No. 5807689
;  GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen
;    APPLICANT:  Pontsler, Aaron
;    APPLICANT:  Johnson, Edwin C.
;    APPLICANT:  Hess, Stephen D.
;    TITLE OF INVENTION:  HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/486,270
;      FILING DATE:  02-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/072,574
;      FILING DATE:  04-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  FP41 9772
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-4737
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  370..3003
;      OTHER INFORMATION:  /product= "HUMAN MGLUR5C"
;      OTHER INFORMATION:  /note= "Variant of MGLUR5A with truncated 3' end."
US-08-486-270-11


  Query Match             56.0%;  Score 14;  DB 1;  Length 3282;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db    2950 TGCTGGATCTGCTG 2937


RESULT  32
US-08-367-264-11/c
; Sequence 11, Application US/08367264
; Patent No. 6001581
;  GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen
;    APPLICANT:  Pontsler, Aaron
;    APPLICANT:  Johnson, Edwin C.
;    APPLICANT:  Hess, Stephen D.
;    TITLE OF INVENTION:  HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/367,264
;      FILING DATE:  02-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/072,574
;      FILING DATE:  04-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  FP41 9772
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-4737
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  370..3003
;      OTHER INFORMATION:  /product= "HUMAN MGLUR5C"
;      OTHER INFORMATION:  /note= "Variant of MGLUR5A with truncated 3' end."
US-08-367-264-11


  Query Match             56.0%;  Score 14;  DB 3;  Length 3282;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db    2950 TGCTGGATCTGCTG 2937


RESULT  33
US-08-072-574-7/c
; Sequence 7, Application US/08072574
; Patent No. 5521297
;  GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen
;    APPLICANT:  Pontsler, Aaron
;    TITLE OF INVENTION:  HUMAN MATABOTROPIC GLUTAMATE RECEPTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  00719
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,574
;      FILING DATE:  19930604
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9383
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4085 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  370..3912
;      OTHER INFORMATION:  /product= "HUMAN MGLUR5A"
US-08-072-574-7


  Query Match             56.0%;  Score 14;  DB 1;  Length 4085;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   2950 TGCTGGATCTGCTG 2937


RESULT  34
US-08-486-270-7/c
; Sequence 7, Application US/08486270
; Patent No. 5807689
;  GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen
;    APPLICANT:  Pontsler, Aaron
;    APPLICANT:  Johnson, Edwin C.
;    APPLICANT:  Hess, Stephen D.
;    TITLE OF INVENTION:  HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/486,270
;      FILING DATE:  02-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/072,574
;      FILING DATE:  04-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  FP41 9772
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-4737
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4085 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  370..3912
;      OTHER INFORMATION:  /product= "HUMAN MGLUR5A"
US-08-486-270-7


  Query Match             56.0%;  Score 14;  DB 1;  Length 4085;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   2950 TGCTGGATCTGCTG 2937


RESULT  35
US-08-367-264-7/c
; Sequence 7, Application US/08367264
; Patent No. 6001581
;  GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen
;    APPLICANT:  Pontsler, Aaron
;    APPLICANT:  Johnson, Edwin C.
;    APPLICANT:  Hess, Stephen D.
;    TITLE OF INVENTION:  HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/367,264
;      FILING DATE:  02-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/072,574
;      FILING DATE:  04-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  FP41 9772
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-4737
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4085 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  both
;       TOPOLOGY:  both
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  370..3912
;       OTHER INFORMATION:  /product= "HUMAN MGLUR5A"
US-08-367-264-7


  Query Match             56.0%;  Score 14;  DB 3;  Length 4085;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   2950 TGCTGGATCTGCTG 2937


RESULT  36
US-08-072-574-9/c
; Sequence 9, Application US/08072574
; Patent No. 5521297
;  GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen
;    APPLICANT:  Pontsler, Aaron
;    TITLE OF INVENTION:  HUMAN MATABOTROPIC GLUTAMATE RECEPTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  00719
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,574
;      FILING DATE:  19930604
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  P41 9383
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  213-622-7700
;      TELEFAX:  213-489-4210
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4181 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  370..4008
;      OTHER INFORMATION:  /product= "HUMAN MGLUR5B"
;      OTHER INFORMATION:  /note= "Variant of MGLUR5A with 96 base pair
;      OTHER INFORMATION:  insertion between nucleotides 2999 and 3000."
US-08-072-574-9


  Query Match             56.0%;  Score 14;  DB 1;  Length 4181;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   2950 TGCTGGATCTGCTG 2937


RESULT  37
US-08-486-270-9/c
; Sequence 9, Application US/08486270
; Patent No. 5807689
;  GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen
;    APPLICANT:  Pontsler, Aaron
;    APPLICANT:  Johnson, Edwin C.
;    APPLICANT:  Hess, Stephen D.
;    TITLE OF INVENTION:  HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/486,270
;      FILING DATE:  02-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/072,574
;      FILING DATE:  04-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  FP41 9772
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-4737
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4181 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  370..4008
;      OTHER INFORMATION:  /product= "HUMAN MGLUR5B"
;      OTHER INFORMATION:  /note= "Variant of MGLUR5A with 96 base pair
;      OTHER INFORMATION:  insertion between nucleotides 2998 and 2999."
US-08-486-270-9


  Query Match             56.0%;  Score 14;  DB 1;  Length 4181;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   2950 TGCTGGATCTGCTG 2937


RESULT  38
US-08-367-264-9/c
```

```
; Sequence 9, Application US/08367264
; Patent No. 6001581
;  GENERAL INFORMATION:
;    APPLICANT:  Daggett, Lorrie
;    APPLICANT:  Ellis, Steven B.
;    APPLICANT:  Liaw, Chen
;    APPLICANT:  Pontsler, Aaron
;    APPLICANT:  Johnson, Edwin C.
;    APPLICANT:  Hess, Stephen D.
;    TITLE OF INVENTION:  HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
;    TITLE OF INVENTION:  NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pretty, Schroeder, Brueggemann & Clark
;      STREET:  444 South Flower Street, Suite 2000
;      CITY:  Los Angeles
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  90071
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/367,264
;      FILING DATE:  02-JUN-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/072,574
;      FILING DATE:  04-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Reiter, Stephen E.
;      REGISTRATION NUMBER:  31,192
;      REFERENCE/DOCKET NUMBER:  FP41 9772
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-546-4737
;      TELEFAX:  619-546-9392
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4181 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  370..4008
;      OTHER INFORMATION:  /product= "HUMAN MGLUR5B"
;      OTHER INFORMATION:  /note= "Variant of MGLUR5A with 96 base pair
;      OTHER INFORMATION:  insertion between nucleotides 2998 and 2999."
US-08-367-264-9


  Query Match             56.0%;  Score 14;  DB 3;  Length 4181;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   2950 TGCTGGATCTGCTG 2937


RESULT  39
US-08-660-148-1/c
; Sequence 1, Application US/08660148
; Patent No. 6211353
;  GENERAL INFORMATION:
;    APPLICANT:  Burnett, J. P.
;    APPLICANT:  Mayne, Nancy G.
;    APPLICANT:  Sharp, Robert L.
;    APPLICANT:  Snyder, Yvonne M.
;    TITLE OF INVENTION:  EXCITATORY AMINO RECEPTOR PROTEIN AND
;    TITLE OF INVENTION:  RELATED NUCLEIC ACID COMPOUNDS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/660,148
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/282,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-9419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4207 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  460..3999
US-08-660-148-1


  Query Match             56.0%;  Score 14;  DB 4;  Length 4207;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   3040 TGCTGGATCTGCTG 3027


RESULT  40
US-08-660-148-3/c
; Sequence 3, Application US/08660148
; Patent No. 6211353
;  GENERAL INFORMATION:
;    APPLICANT:  Burnett, J. P.
;    APPLICANT:  Mayne, Nancy G.
;    APPLICANT:  Sharp, Robert L.
;    APPLICANT:  Snyder, Yvonne M.
;    TITLE OF INVENTION:  EXCITATORY AMINO RECEPTOR PROTEIN AND
;    TITLE OF INVENTION:  RELATED NUCLEIC ACID COMPOUNDS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/660,148
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/282,853
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Gaylo, Paul J.
;       REGISTRATION NUMBER:  36,808
;       REFERENCE/DOCKET NUMBER:  X-9419
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (317) 276-0756
;       TELEFAX:  (317) 276-3861
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4207 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  mRNA
US-08-660-148-3


  Query Match             56.0%;  Score 14;  DB 4;  Length 4207;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db    3040 TGCTGGATCTGCTG 3027


RESULT  41
US-08-660-148-4/c
; Sequence 4, Application US/08660148
; Patent No. 6211353
;  GENERAL INFORMATION:
;    APPLICANT:  Burnett, J. P.
;    APPLICANT:  Mayne, Nancy G.
;    APPLICANT:  Sharp, Robert L.
;    APPLICANT:  Snyder, Yvonne M.
;    TITLE OF INVENTION:  EXCITATORY AMINO RECEPTOR PROTEIN AND
;    TITLE OF INVENTION:  RELATED NUCLEIC ACID COMPOUNDS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/660,148
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/282,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-9419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4303 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  460..4095
US-08-660-148-4


  Query Match             56.0%;  Score 14;  DB 4;  Length 4303;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctggatctgctg 16
           ||||||||||||||
Db    3040 TGCTGGATCTGCTG 3027


RESULT  42
US-08-660-148-6/c
; Sequence 6, Application US/08660148
; Patent No. 6211353
;  GENERAL INFORMATION:
;    APPLICANT:  Burnett, J. P.
;    APPLICANT:  Mayne, Nancy G.
;    APPLICANT:  Sharp, Robert L.
;    APPLICANT:  Snyder, Yvonne M.
;    TITLE OF INVENTION:  EXCITATORY AMINO RECEPTOR PROTEIN AND
;    TITLE OF INVENTION:  RELATED NUCLEIC ACID COMPOUNDS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/660,148
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/282,853
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-9419
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4303 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  mRNA
US-08-660-148-6


  Query Match             56.0%;  Score 14;  DB 4;  Length 4303;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      3 tgctggatctgctg 16
          ||||||||||||||
Db   3040 TGCTGGATCTGCTG 3027


RESULT  43
US-08-465-976A-1/c
; Sequence 1, Application US/08465976A
; Patent No. 5869632
;  GENERAL INFORMATION:
;    APPLICANT:  SOPPET, DANIEL R
;    APPLICANT:  LI, YI
;    APPLICANT:  ROSEN, CRAIG A
;    APPLICANT:  RUBEN, STEVEN M
;    TITLE OF INVENTION:  HUMAN G-PROTEIN RECEPTOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CARELLA, BYRNE, BAIN GILFILLAN, CECCHI
;      ADDRESSEE:  STEWART & OLSTEIN
;      STREET:  6 BECKER FARM ROAD
;      CITY:  ROSELAND
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07068
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,976A
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FERRARO, GREGORY F
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-444
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201) 994-1700
;      TELEFAX:  (201) 994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4566 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  212..2863
US-08-465-976A-1


  Query Match           56.0%;  Score 14;  DB 2;  Length 4566;
  Best Local Similarity  100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 ctggatctgctgtt 18
          ||||||||||||||
Db   4182 CTGGATCTGCTGTT 4169


RESULT  44
US-08-982-412-1/c
; Sequence 1, Application US/08982412
; Patent No. 5958729
;  GENERAL INFORMATION:
;    APPLICANT:  SOPPET, DANIEL R
;    APPLICANT:  LI, YI
;    APPLICANT:  ROSEN, CRAIG A
;    APPLICANT:  RUBEN, STEVEN M
;    TITLE OF INVENTION:  HUMAN G-PROTEIN RECEPTOR
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HUMAN GENOME SCIENCES, INC.
;      STREET:  9410 KEY WEST AVENUE
;      CITY:  ROCKVILLE,
;      STATE:  MD
;      COUNTRY:  US
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/982,412
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BROOKES, ANDERS A
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PF181PCT2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4566 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  212..2863
US-08-982-412-1


  Query Match           56.0%;  Score 14;  DB 2;  Length 4566;
  Best Local Similarity  100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 ctggatctgctgtt 18
          ||||||||||||||
Db   4182 CTGGATCTGCTGTT 4169


RESULT  45
US-08-169-927-1
; Sequence 1, Application US/08169927
; Patent No. 5783441
;  GENERAL INFORMATION:
;    APPLICANT:  Carl, Mitchell
;    APPLICANT:  Dobson, Michael E.
;    APPLICANT:  Ching, Wei Mei
;    APPLICANT:  Dasch, Gregory A
;    TITLE OF INVENTION:  Gene and Protein Applicable to the
;    TITLE OF INVENTION:  Preparation of Vaccines for Rickettsia prowazekii and
;    TITLE OF INVENTION:  Rickettsia typhi and the Detection of Both
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Counsel, Naval Medical R & D Command
;      STREET:  Bldg. 1, T-12, 8901 Wisconsin Ave.
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20889-5606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/169,927
;      FILING DATE:
```

```
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/742,128
;      FILING DATE:  08/09/91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spevack, A. David
;      REGISTRATION NUMBER:  24,743
;      REFERENCE/DOCKET NUMBER:  75,976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-1022
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5319 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
;      ORGANISM:  Rickettsia prowazekii
;      STRAIN:  Breinl
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  340..345
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  363..368
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  391..5226
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  379..386
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5270..5306
;    PUBLICATION INFORMATION:
;      AUTHORS:  Carl, M.
;      AUTHORS:  Dobson, M. E.
;      AUTHORS:  Ching, W. M.
;      AUTHORS:  Dasch, G. A.
;      TITLE:  Characterization of the gene encoding the
;      TITLE:  protective S-layer protein of Rickettsia
;      TITLE:  prowazekii; presence of a truncated identical
;      TITLE:  homolog in rickettsia typhi
;      JOURNAL:  Proc. Natl. Acad. Sci. U.S.A.
;      DATE:  1990
;      RELEVANT RESIDUES IN SEQ ID NO:  1: FROM 1 TO 5319
US-08-169-927-1


  Query Match             56.0%;  Score 14;  DB 1;  Length 5319;
  Best Local Similarity   100.0%;  Pred. No. 25;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgc 14
           ||||||||||||||
Db    4321 GTTGCTGGATCTGC 4334


Search completed: December 26, 2001, 12:52:35
Job time: 6787 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:09:56 ; Search time 4619.78 Seconds
(without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196F-3
Perfect score: 25
Sequence: 1 gttgctggatctgctgtttgaagcg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 9

Total number of hits satisfying chosen parameters: 862398

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19	76.0	888	13	CNS04KH2	AL294815 Tetraodon
c	2	18	72.0	403	10	BE095496	BE095496 UI-R-BU0-
	3	18	72.0	487	11	BF428015	BF428015 daa04g08.
c	4	17	68.0	546	13	AQ952526	AQ952526 Sheared D
c	5	17	68.0	547	10	AV718673	AV718673 AV718673
	6	16	64.0	237	10	AW101291	AW101291 sd77b02.y
	7	16	64.0	247	10	BE023256	BE023256 sm80a02.y
c	8	16	64.0	274	11	BF414722	BF414722 UI-R-BJ2-
	9	16	64.0	287	10	BB455974	BB455974 BB455974
	10	16	64.0	287	11	BI059872	BI059872 IL3-UT011
c	11	16	64.0	287	11	BF917535	BF917535 IL3-UT011
	12	16	64.0	293	13	AZ214122	AZ214122 Sheared D
	13	16	64.0	301	10	AW306120	AW306120 se46b06.y
c	14	16	64.0	328	10	BE659333	BE659333 GM700009A
	15	16	64.0	342	10	AW423279	AW423279 sh65d11.y
c	16	16	64.0	360	11	C61070	C61070 C61070 Yuji
c	17	16	64.0	365	10	BE662000	BE662000 bs02d05.y
c	18	16	64.0	412	11	BI054795	BI054795 MR3-GN034
	19	16	64.0	422	10	AI736967	AI736967 sb35c02.y
	20	16	64.0	434	10	AW704839	AW704839 sk40b11.y
c	21	16	64.0	436	10	AI293622	AI293622 LP06802.5
	22	16	64.0	441	10	BE190258	BE190258 so09f08.y
	23	16	64.0	470	10	AL373871	AL373871 MtBB03D01
	24	16	64.0	474	10	AW704996	AW704996 sk41c02.y
	25	16	64.0	481	10	AL379973	AL379973 MtBB48E08
c	26	16	64.0	497	11	BF461669	BF461669 UI-M-CG0p
	27	16	64.0	502	10	BE515828	BE515828 WHE0606_B
	28	16	64.0	503	10	AL383090	AL383090 MtBC11G02
	29	16	64.0	513	11	BI418981	BI418981 LjNEST51h
	30	16	64.0	517	10	AL373872	AL373872 MtBB03D01
	31	16	64.0	517	13	AQ692054	AQ692054 HS_5408_A
c	32	16	64.0	521	11	BF419551	BF419551 UI-R-CA0-
c	33	16	64.0	522	10	AI854720	AI854720 UI-M-BH0-
	34	16	64.0	525	10	AL379974	AL379974 MtBB48E08
	35	16	64.0	531	10	AW395303	AW395303 sh46d09.y
c	36	16	64.0	555	13	AQ889100	AQ889100 HS_2204_A
c	37	16	64.0	570	11	BI229421	BI229421 RE27194.5
	38	16	64.0	574	10	AW573807	AW573807 EST316398
	39	16	64.0	575	11	BE999070	BE999070 EST430793
	40	16	64.0	580	11	BF525150	BF525150 UI-R-AF0-
	41	16	64.0	591	10	AW734894	AW734894 sk76d03.y
	42	16	64.0	601	13	AZ267971	AZ267971 RPCI-23-1
c	43	16	64.0	602	11	BI229406	BI229406 RE27178.5
	44	16	64.0	609	10	AW267738	AW267738 EST305866
	45	16	64.0	616	10	AW980891	AW980891 EST392044
	46	16	64.0	618	11	BG585328	BG585328 EST487092
c	47	16	64.0	621	10	AW422111	AW422111 fi57a12.y
c	48	16	64.0	627	10	BE593142	BE593142 WS1_98_A0
	49	16	64.0	632	13	AZ449344	AZ449344 1M0247J04
c	50	16	64.0	634	11	BG634485	BG634485 AT30708.5
c	51	16	64.0	640	10	AI108765	AI108765 GH08003.5
	52	16	64.0	643	13	AZ015305	AZ015305 RPCI-23-3
c	53	16	64.0	650	13	TA69C05P	AL457556 T. brucei
c	54	16	64.0	755	10	BE034735	BE034735 ML03G01 M
c	55	16	64.0	831	11	BG584066	BG584066 EST485826
c	56	16	64.0	834	13	CNS01VI3	AL169140 Tetraodon
	57	16	64.0	1104	13	CNS02ARM	AL188923 Tetraodon
c	58	15	60.0	161	11	BI389308	BI389308 949052E12
	59	15	60.0	177	13	FR0044714	AL132206 Fugu rubr
	60	15	60.0	190	11	BI024308	BI024308 CM3-MT029
c	61	15	60.0	196	10	BB398642	BB398642 BB398642
c	62	15	60.0	227	10	BE012248	BE012248 945016A03
c	63	15	60.0	243	10	AI547503	AI547503 UI-R-C3-s
c	64	15	60.0	277	11	BE980507	BE980507 UI-M-BG2-
c	65	15	60.0	278	10	AI021028	AI021028 ua99b01.r
	66	15	60.0	295	11	BF440174	BF440174 BS2900018
	67	15	60.0	298	10	BB337949	BB337949 BB337949
c	68	15	60.0	300	11	C35496	C35496 C35496 Yuji
c	69	15	60.0	300	11	C51717	C51717 C51717 Yuji
c	70	15	60.0	300	11	C52955	C52955 C52955 Yuji
c	71	15	60.0	300	11	C59449	C59449 C59449 Yuji
	72	15	60.0	301	11	BE990261	BE990261 UI-M-BZ1-
	73	15	60.0	303	11	BE990318	BE990318 UI-M-BZ1-
c	74	15	60.0	316	13	FR0039786	AL127281 Fugu rubr
	75	15	60.0	342	10	AV209024	AV209024 AV209024
c	76	15	60.0	348	13	AZ228352	AZ228352 RPCI-23-7
c	77	15	60.0	356	11	BG375907	BG375907 UI-R-CU0-
c	78	15	60.0	358	11	BF355075	BF355075 RC2-HT082
	79	15	60.0	360	10	AI486110	AI486110 EST244431
	80	15	60.0	372	13	AQ521765	AQ521765 HS_5239_A
c	81	15	60.0	398	11	BI389121	BI389121 949052E12
	82	15	60.0	403	10	AA677232	AA677232 zj55h04.s
c	83	15	60.0	406	13	AQ143500	AQ143500 HS_3066_B
	84	15	60.0	424	13	AQ494880	AQ494880 HS_5195_A
	85	15	60.0	427	13	BH077581	BH077581 RPCI-24-2

86	15	60.0	435	13	AQ114851	AQ114851 CIT-HSP-2
87	15	60.0	437	10	AW432846	AW432846 sh81g05.y
88	15	60.0	442	10	AA397373	AA397373 nc64c07.s
c 89	15	60.0	443	11	BF415104	BF415104 UI-R-BJ2-
90	15	60.0	445	10	BE580608	BE580608 kp79d11.y
c 91	15	60.0	456	11	BF555582	BF555582 UI-R-E0-c
c 92	15	60.0	458	11	BG379251	BG379251 UI-R-BT1-
93	15	60.0	479	13	AZ965014	AZ965014 2M0234K24
94	15	60.0	483	10	AI958363	AI958363 fd03g04.y
95	15	60.0	489	13	CNS04S8M	AL304879 Tetraodon
96	15	60.0	495	13	AQ192174	AQ192174 HS_3227_B
97	15	60.0	501	10	AI984034	AI984034 wt52f02.x
c 98	15	60.0	502	10	AW660235	AW660235 98156_MAR
c 99	15	60.0	507	10	BE114310	BE114310 UI-R-BJ1-
c 100	15	60.0	509	10	AW915326	AW915326 EST346630

ALIGNMENTS

RESULT 1
CNS04KH2
LOCUS CNS04KH2 888 bp DNA GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 116D03 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL294815
VERSION AL294815.1 GI:8033395
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 888)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 888)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 888)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .888
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="116D03"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG116CB02LP1-end : T7"
BASE COUNT 129 a 276 c 276 g 187 t 20 others
ORIGIN

Query Match 76.0%; Score 19; DB 13; Length 888;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgctggatctgctgtttga 21
||||||||||||||||||||
Db 331 TGCTGGATCTGCTGTTTGA 349

RESULT 2
BE095496/c
LOCUS BE095496 403 bp mRNA EST 12-JUN-2000
DEFINITION UI-R-BU0-apa-c-11-0-UI.s1 UI-R-BU0 Rattus norvegicus cDNA clone UI-R-BU0-apa-c-11-0-UI 3', mRNA sequence.
ACCESSION BE095496
VERSION BE095496.1 GI:8486427
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 403)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=No.
FEATURES Location/Qualifiers
source 1. .403
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BU0-apa-c-11-0-UI"
/clone_lib="UI-R-BU0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BU0 library is a subtracted library derived from a mixture of eye and ganglia tissues. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_SEQ=None found"
BASE COUNT 85 a 147 c 141 g 30 t
ORIGIN

Query Match 72.0%; Score 18; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctgtt 18
||||||||||||||||||
Db 245 GTTGCTGGATCTGCTGTT 228

RESULT 3
BF428015
LOCUS BF428015 487 bp mRNA EST 30-MAR-2001
DEFINITION daa04g08.x1 NICHD XGC Ov1 Xenopus laevis cDNA clone IMAGE:4055247 3' similar to TR:O93478 O93478 KINESIN LIKE PROTEIN 3. ;, mRNA sequence.

```
ACCESSION   BF428015
VERSION     BF428015.1  GI:11439676
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 487)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 50.
FEATURES             Location/Qualifiers
     source          1. .487
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4055247"
                     /clone_lib="NICHD XGC Ov1"
                     /sex="female"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: Not1;
                     Site_2: Sal1; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.0 kb. Constructed by Life
                     Technologies."
BASE COUNT      138 a    101 c    102 g    146 t
ORIGIN


  Query Match             72.0%;  Score 18;  DB 11;  Length 487;
  Best Local Similarity   100.0%;  Pred. No. 9.4;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgtt 18
          ||||||||||||||||||
Db    249 GTTGCTGGATCTGCTGTT 266


RESULT   4
AQ952526/c
LOCUS       AQ952526      546 bp    DNA             GSS       27-JAN-2000
DEFINITION  Sheared DNA-48D19.TF Sheared DNA Trypanosoma brucei genomic clone
            Sheared DNA-48D19, DNA sequence.
ACCESSION   AQ952526
VERSION     AQ952526.1  GI:6775791
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
            Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.
            , Fraser,C. and Adams,M.
  TITLE     Determination of clone end sequences from Trypanosoma brucei GUTat
            10.1 sheared DNA library
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through ATCC. Sheared DNA end sequences search page:
            http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Trypanosoma brucei"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-48D19"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
BASE COUNT      192 a    161 c    127 g     66 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 13;  Length 546;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctgt 17
          |||||||||||||||||
Db    192 GTTGCTGGATCTGCTGT 176


RESULT   5
AV718673/c
LOCUS       AV718673      547 bp    mRNA            EST       16-OCT-2000
DEFINITION  AV718673 GLC Homo sapiens cDNA clone GLCDYA12 5', mRNA sequence.
ACCESSION   AV718673
VERSION     AV718673.1  GI:10815825
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.
            , Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
            Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang
            ,Y., Gu,Y., Chen,Z. and Han,Z.
  TITLE     Homo sapiens cDNA GLC clones
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Zequang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GLCDYA12"
                     /clone_lib="GLC"
                     /tissue_type="corresponding non cancerous liver tissue"
```

```
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       183 a    101 c    116 g    147 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 10;  Length 547;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctgt 17
           |||||||||||||||||
Db     473 GTTGCTGGATCTGCTGT 457


RESULT   6
AW101291
LOCUS       AW101291      237 bp    mRNA            EST       06-DEC-1999
DEFINITION  sd77b02.y1 Gm-c1009 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1009-388 5' similar to TR:Q40380 Q40380 ARABINOGALACTAN-PROTEIN
            PRECURSOR. ;, mRNA sequence.
ACCESSION   AW101291
VERSION     AW101291.1  GI:6071904
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 237)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .237
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1009-388"
                     /clone_lib="Gm-c1009"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; The mRNA was isolated from entire roots of
                     2-month-old 'Williams' plants that were greenhouse grown
                     in 5-gallon pots. To suppress nodulation, Black Gold
                     All-Purpose potting soil was supplemented with: 0.36g/L
                     available phosphoric acid (P2O5), 20mg/L urea N, 0.16g/L
                     S, 0.49mg/L B, 2.5mg/L Cu, 0.15g/L Fe, 13.53mg/L Mn,
                     0.26mg/L Mo, 14mg/L Zn, 20mg/L Ca, and the following
                     nutrients in a slow-release form (Osmocote): 0.165g/L
                     ammonia N, 0.185g/L nitrate N, 0.35g/L available
                     phosphoric acid, and 0.35g/L soluble potash. No nodules
                     were visible on the roots at harvest. Stratagene's cDNA
                     Synthesis Kit (catalog #200401) was used to synthesize the
                     cDNA. First-strand synthesis was performed with 5-methyl
                     dCTP, hence the ligated cDNA is hemimethylated.
                     Stratagene's first-strand synthesis primer was used
                     [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)-18]. After
                     second-strand synthes is, the cDNA ends were 'polished'
                     with clone Pfu DNA polymerase, ligated to EcoRI adapters,
                     and phosphorylated. The XhoI site within the first-strand
                     synthesis primer was restricted by digestion with XhoI;
                     all XhoI sites in the cDNA would be protected by their
                     hemimethylated status. The cDNA constructs were
                     size-fractionated with a 400bp cutoff, using a SizeSep 400
                     Spun column from Pharmacia. The column eluent was then
                     ligated into Stratagene's pBluescript II XR Predigested
                     vector (pBluescript II SK(+) that had been digested with
                     EcoRI and XhoI, and phosphorylated). Both the white and
                     blue colonies appear to contain recombinant plasmids with
                     cDNA inserts. This library was constructed by Dr. Paul
                     Keim and Dr. Virginia Coryell."
BASE COUNT       40 a     60 c     52 g     85 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 237;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db      96 GTTGCTGGATCTGCTG 111


RESULT   7
BE023256
LOCUS       BE023256      247 bp    mRNA            EST       21-NOV-2000
DEFINITION  sm80a02.y1 Gm-c1015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1015-6339 5', mRNA sequence.
ACCESSION   BE023256
VERSION     BE023256.1  GI:8285697
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 247)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Insert Length: 438   Std Error: 0.00
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .247
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
```

/clone="GENOME SYSTEMS CLONE ID: Gm-c1015-6339"
/clone_lib="Gm-c1015"
/tissue_type="Mature flowers, field grown plants"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."

BASE COUNT 39 a 65 c 57 g 77 t 9 others
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 247;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 128 GTTGCTGGATCTGCTG 143

RESULT 8
BF414722/c
LOCUS BF414722 274 bp mRNA EST 28-NOV-2000
DEFINITION UI-R-BJ2-bow-c-11-0-UI.s1 UI-R-BJ2 Rattus norvegicus cDNA clone UI-R-BJ2-bow-c-11-0-UI 3', mRNA sequence.
ACCESSION BF414722
VERSION BF414722.1 GI:11402711
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 274)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=No.
FEATURES Location/Qualifiers
source 1. .274
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ2-bow-c-11-0-UI"
/clone_lib="UI-R-BJ2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2 library is a subtracted library derived from the following tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc, atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15 dpc, AV canal at 15 dpc. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996) TAG_SEQ=None found"

BASE COUNT 54 a 96 c 106 g 18 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 274;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 246 GTTGCTGGATCTGCTG 231

RESULT 9
BB455974
LOCUS BB455974 287 bp mRNA EST 21-JUL-2000
DEFINITION BB455974 RIKEN full-length enriched, 12 days embryo spinal ganglion Mus musculus cDNA clone D130043H15 3', mRNA sequence.
ACCESSION BB455974
VERSION BB455974.1 GI:9351467
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 287)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1. .287
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="D130043H15"
                     /clone_lib="RIKEN full-length enriched, 12 days embryo
                     spinal ganglion"
                     /tissue_type="spinal ganglion"
                     /dev_stage="12 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT       97 a     48 c     33 g    109 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 287;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 ttgctggatctgctgt 17
           ||||||||||||||||
Db      93 TTGCTGGATCTGCTGT 108


RESULT  10
BI059872
LOCUS       BI059872      287 bp    mRNA            EST       15-JUN-2001
DEFINITION  IL3-UT0115-080101-393-F03 UT0115 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI059872
VERSION     BI059872.1  GI:14467399
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&t2=IL3-UT0115-
            080101-393-F03&t3=2001-01-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 287.
FEATURES             Location/Qualifiers
     source          1. .287
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="UT0115"
                     /dev_stage="Adult"
                     /note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       65 a     97 c     62 g     63 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 287;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db     197 GTTGCTGGATCTGCTG 212


RESULT  11
BF917535/c
LOCUS       BF917535      287 bp    mRNA            EST       18-JAN-2001
DEFINITION  IL3-UT0115-201200-392-D09 UT0115 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF917535
VERSION     BF917535.1  GI:12308993
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 287)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&t2=IL3-UT0115-
            201200-392-D09&t3=2000-12-20&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 287.
FEATURES             Location/Qualifiers
     source          1. .287
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="UT0115"
                     /dev_stage="Adult"
                     /note="Organ: uterus_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
```

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 63 a 62 c 97 g 65 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 287;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 91 GTTGCTGGATCTGCTG 76

RESULT 12
AZ214122
LOCUS AZ214122 293 bp DNA GSS 09-JUN-2000
DEFINITION Sheared DNA-67G1.TF Sheared DNA Trypanosoma brucei genomic clone Sheared DNA-67G1, DNA sequence.
ACCESSION AZ214122
VERSION AZ214122.1 GI:8431922
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 293)
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: Sheared DNA-67G1.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through Research Genetics, Alabama, USA. Sheared DNA end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/.
Seq primer: M13-Forward
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .293
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-67G1"
/clone_lib="Sheared DNA"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
BASE COUNT 38 a 73 c 78 g 104 t
ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 293;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atctgctgtttgaagc 24
||||||||||||||||
Db 149 ATCTGCTGTTTGAAGC 164

RESULT 13
AW306120
LOCUS AW306120 301 bp mRNA EST 20-JAN-2000
DEFINITION se46b06.y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1017-1932 5', mRNA sequence.
ACCESSION AW306120
VERSION AW306120.1 GI:6718473
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 301)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco.
FEATURES Location/Qualifiers
source 1. .301
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1017-1932"
/clone_lib="Gm-c1017"
/tissue_type="vegetable buds of field grown plants"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from vegatable buds of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."
BASE COUNT 73 a 48 c 60 g 119 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 301;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db     45 GTTGCTGGATCTGCTG 60


RESULT  14
BE659333/c
LOCUS       BE659333      328 bp    mRNA            EST       24-MAY-2001
DEFINITION  GM700009A20A6 Gm-r1070 Glycine max cDNA clone Gm-r1070-3084 3',
            mRNA sequence.
ACCESSION   BE659333
VERSION     BE659333.1  GI:9985225
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 328)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AW423279 corresponding to Gm-c1015-4654 (5')
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES             Location/Qualifiers
     source          1. .328
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1070-3084"
                     /clone_lib="Gm-r1070"
                     /note="The library Gm-r1070 is a sequence-driven, reracked
                     set of 9,216  clones selected from cDNA libraries from
                     various tissues and stages  of development of soybean that
                     represent 2,639 sequences from  immature cotyledons, 1,770
                     from immature seed coats, 3,938 from  flowers, and 869
                     from young pods.  The 5' ESTs of the source clones from
                     the different libraries was  used to select singletons, or
                     a representative of each contig, which  were reracked to
                     form library Gm-r1070.  The cDNA clones of the reracked
                     Gm-r1070 library were then sequenced  at the 3' end.  The
                     contig analysis to select unique genes was performed by
                     the  laboratory of Ernest Retzel, Center for Computational
                     Genomics and  Bioinformatics, University of Minnesota,
                     http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                     .  Reracking was performed by Genome Systems, St. Louis,
                     http://www.genomesystems.com, and 3' sequencing by the
                     Keck Center  for Comparative and Functional Genomics,
                     University of Illinois,
                     http://www.life.uiuc.edu/biotech/keck.html.  Note: The
                     corresponding 5' EST from each clone in the Gm-r1070
                     library is listed in the 'OTHER EST' field.  The detailed
                     information on the source library for each clone can  also
                     be obtained by referring to the Genome Systems clone ID o
                     the  original cDNA library that is also listed under
                     'OTHER EST'."
BASE COUNT      115 a     71 c     76 g     53 t     13 others
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 10;  Length 328;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db    226 GTTGCTGGATCTGCTG 211


RESULT  15
AW423279
LOCUS       AW423279      342 bp    mRNA            EST       09-FEB-2000
DEFINITION  sh65d11.y1 Gm-c1015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1015-4654 5' similar to TR:Q40380 Q40380
            ARABINOGALACTAN-PROTEIN PRECURSOR. ;, mRNA sequence.
ACCESSION   AW423279
VERSION     AW423279.1  GI:6951280
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 342)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 328.
FEATURES             Location/Qualifiers
     source          1. .342
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1015-4654"
                     /clone_lib="Gm-c1015"
                     /tissue_type="Mature flowers, field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from mature flowers of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR cDNA library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT       70 a     72 c     70 g    130 t
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 10;  Length 342;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db    101 GTTGCTGGATCTGCTG 116


RESULT  16
C61070/c
LOCUS       C61070          360 bp    mRNA            EST       22-SEP-1997
DEFINITION  C61070 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk224c7 5', mRNA sequence.
ACCESSION   C61070
VERSION     C61070.1  GI:2419775
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
            ,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk224c7"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT      101 a     71 c     76 g    111 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 atctgctgtttgaagc 24
          ||||||||||||||||
Db     26 ATCTGCTGTTTGAAGC 11


RESULT  17
BE662000/c
LOCUS       BE662000        365 bp    mRNA            EST       08-JAN-2001
DEFINITION  bs02d05.y1 Drosophila melanogaster adult testis library Drosophila
            melanogaster cDNA clone bs02d05 5', mRNA sequence.
ACCESSION   BE662000
VERSION     BE662000.1  GI:9992572
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 365)
  AUTHORS   Andrews,J., Bouffard,G.G., Cheadle,C., Lu,J., Becker,K.G. and
            Oliver,B.
  TITLE     Gene discovery using computational and microarray analysis of
            transcription in the drosophila melanogaster testis
  JOURNAL   Genome Res. 10 (12), 2030-2043 (2000)
  MEDLINE   20568492
COMMENT     Contact: Brian Oliver
            Laboratory of Cellular and Developmental Biology
            NIDDK, National Institutes of Health
            6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
            Fax: (301) 496 5239
            Email: oliver@helix.nih.gov,
            http://www.niddk.nih.gov/intram/people/boliver.htm
            Tissue isolation and library construction performed at the National
            Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
            http://www.niddk.nih.gov/intram/people/boliver.htm). DNA sequencing
            and analyses performed by National Institutes of Health Intramural
            Sequencing Center (NISC; see http://www.nisc.nih.gov).
            Plate: 02  row: d  column: 05
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .365
                     /organism="Drosophila melanogaster"
                     /strain="y[*] w[67c1]/Y"
                     /db_xref="taxon:7227"
                     /clone="bs02d05"
                     /clone_lib="Drosophila melanogaster adult testis library"
                     /sex="male"
                     /dev_stage="1-5 day adult"
                     /lab_host="SOLR (Stratagene)"
                     /note="Organ: testis; Vector: pBlueScript SK (Stratagene);
                     Site_1: EcoR I; Site_2: Xho I; Testes dissected from 1-5
                     day adult y[*] w[67c1]/Y males raised at 25oC.  RNA
                     isolated using Trizol (Life Technologies) and a single
                     round of Poly(A)+ selection using Oligotex (Qiagen).  cDNA
                     library constructed using Stratagene ZAP-cDNA syntesis
                     kit.  Oligo dT-primed, size fractionated ~1-6 kb, and
                     directionally cloned at EcoRI and XhoI in Uni-ZAP XR.
                     Following a single round of amplification pBlueScript SK
                     phagemids were mass excised.  A distribution channel for
                     clones is being sought, but not currently available.
                     Requests for clones cannot be honored."
BASE COUNT      85 a    129 c     94 g     56 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 365;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db    330 GTTGCTGGATCTGCTG 315


RESULT  18
BI054795/c
LOCUS       BI054795        412 bp    mRNA            EST       15-JUN-2001
DEFINITION  MR3-GN0349-020201-003-f12 GN0349 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI054795
VERSION     BI054795.1  GI:14462325
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
```

```
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=MR3&t2=MR3-GN0349-
            020201-003-f12&t3=2001-02-02&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .412
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="GN0349"
                     /dev_stage="Adult"
                     /note="Organ: placenta_normal; Vector: puc18; Site_1: SmaI
                     ; Site_2: SmaI; A mini-library was made by cloning
                     products derived from  ORESTES PCR (U.S. Letters Patent
                     application No. 196,716 -  Ludwig Institute for Cancer
                     Research) profiles into  the pUC 18 vector. Reverse
                     transcription of tissue mRNA and cDNA amplification were
                     performed under low stringency conditions."
BASE COUNT      102 a    117 c    107 g     86 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 412;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db     397 GTTGCTGGATCTGCTG 382


RESULT  19
AI736967
LOCUS       AI736967      422 bp    mRNA            EST       31-JUL-2001
DEFINITION  sb35c02.y1 Gm-c1013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1013-27 5' similar to TR:Q40380 Q40380 ARABINOGALACTAN-PROTEIN
            PRECURSOR. ;, mRNA sequence.
ACCESSION   AI736967
VERSION     AI736967.1  GI:5058435
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Insert Length: 530    Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 409.
FEATURES             Location/Qualifiers
     source          1. .422
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1013-27"
                     /clone_lib="Gm-c1013"
                     /tissue_type="Whole seedlings, 2-3 week old seedlings,
                     greenhouse grown"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from whole seedlings of 2-3 week old greenhouse grown
                     plants. The cDNA library was prepared using the Stratagene
                     pBluescript II XR cDNA library construction kit.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting of a poly (dT) sequence with a XhoI restriction
                     site. EcoRI adapters were ligated to the blunt-ended cDNA
                     fragments followed by XhoI digestion. The cDNA fragments
                     were directionally cloned into the EcoRI-XhoI restriction
                     site of the pBluescript vector. The ligated cDNA fragments
                     were transformed into XL10-Gold host cells. This library
                     was constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT       83 a    105 c     91 g    138 t      5 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 422;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db     183 GTTGCTGGATCTGCTG 198


RESULT  20
AW704839
LOCUS       AW704839      434 bp    mRNA            EST       21-NOV-2000
DEFINITION  sk40b11.y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1019-5014 5', mRNA sequence.
ACCESSION   AW704839
VERSION     AW704839.1  GI:7589057
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
```

```
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Insert Length: 553   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 413.
FEATURES              Location/Qualifiers
     source           1. .434
                      /organism="Glycine max"
                      /db_xref="taxon:3847"
                      /clone="GENOME SYSTEMS CLONE ID: Gm-c1019-5014"
                      /clone_lib="Gm-c1019"
                      /tissue_type="immature seed coats of greenhouse grown
                      plants"
                      /lab_host="DH10B (Gibco BRL)"
                      /note="Vector: pSPORT1 (Life Technologies); Site_1: Not I;
                      Site_2: Sal I; This cDNA library was constructed from mRNA
                      isolated from immature seed coats (200-300 mgs) of
                      greenhouse grown plants. The library was prepared using
                      the Life Technologies pSuperScript cDNA library
                      construction kit. Complementary DNA was synthesized from
                      mRNA using a poly (dT) sequence with a Not I restriction
                      site. Sal I linker adapters were ligated to the
                      blunt-ended cDNA fragments followed by Not I digestion.
                      The cDNA fragments were directionally cloned into the Not
                      I-Sal I restriction site of the pSPORT1 vector. The
                      ligated cDNA fragments were transformed into E.coli
                      ElectroMax DH10B host cells (Gibco BRL). This library was
                      constructed by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT       57 a    169 c     90 g    118 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 434;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db    398 GTTGCTGGATCTGCTG 413


RESULT  21
AI293622/c
LOCUS       AI293622      436 bp    mRNA            EST       19-APR-2001
DEFINITION  LP06802.5prime LP Drosophila melanogaster larval-early pupal pOT2
            Drosophila melanogaster cDNA clone LP06802 5prime, mRNA sequence.
ACCESSION   AI293622
VERSION     AI293622.1  GI:3943029
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 436)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 68  row: A  column: 2
            High quality sequence stop: 390.
FEATURES              Location/Qualifiers
     source           1. .436
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="LP06802"
                      /clone_lib="LP Drosophila melanogaster larval-early pupal
                      pOT2"
                      /sex="male and female"
                      /dev_stage="larvae-pupae"
                      /lab_host="DH5-alpha"
                      /note="Organ: whole body; Vector: pOT2; Site_1: EcoRI;
                      Site_2: XhoI; Sized fractionated cDNAs were directly
                      ligated into pOT2. Plasmid cDNA library.
                      "
BASE COUNT      111 a    122 c    111 g     92 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 436;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db    328 GTTGCTGGATCTGCTG 313


RESULT  22
BE190258
LOCUS       BE190258      441 bp    mRNA            EST       22-JUN-2000
DEFINITION  so09f08.y1 Gm-c1035 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1035-2848 5', mRNA sequence.
ACCESSION   BE190258
VERSION     BE190258.1  GI:8669151
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 441)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            High quality sequence stop: 440.
FEATURES              Location/Qualifiers
     source           1. .441
                      /organism="Glycine max"
                      /db_xref="taxon:3847"
                      /clone="GENOME SYSTEMS CLONE ID: Gm-c1035-2848"
                      /clone_lib="Gm-c1035"
                      /tissue_type="immature leaves of greenhouse grown plants"
                      /lab_host="DH10B"
                      /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
                      cDNA library was constructed from mRNA isolated from
                      immature leaves (unfurled trifoliate) of greenhouse grown
```

plants that were 2 weeks old. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a NotI restrictions site. SalI linkers adapters were ligated to the blunt-ended cDNA fragments followed by NotI digestion. The cDNA fragments were directionally cloned into the NotI-SalI restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"
BASE COUNT 102 a 63 c 85 g 190 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 441;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 48 GTTGCTGGATCTGCTG 63

RESULT 23
AL373871
LOCUS AL373871 470 bp mRNA EST 03-AUG-2000
DEFINITION MtBB03D01F1 MtBB Medicago truncatula cDNA clone MtBB03D01 T3, mRNA sequence.
ACCESSION AL373871
VERSION AL373871.1 GI:9673623
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 470)
AUTHORS Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V. and Gamas,P.
TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules
JOURNAL Unpublished (2000)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES Location/Qualifiers
source 1. .470
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MtBB03D01"
/clone_lib="MtBB"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
/note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."
BASE COUNT 85 a 137 c 68 g 180 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 470;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 249 GTTGCTGGATCTGCTG 264

RESULT 24
AW704996
LOCUS AW704996 474 bp mRNA EST 18-APR-2000
DEFINITION sk41c02.y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1019-5115 5', mRNA sequence.
ACCESSION AW704996
VERSION AW704996.1 GI:7589217
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 474)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 434.
FEATURES Location/Qualifiers
source 1. .474
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1019-5115"
/clone_lib="Gm-c1019"
/tissue_type="immature seed coats of greenhouse grown plants"
/lab_host="DH10B (Gibco BRL)"
/note="Vector: pSPORT1 (Life Technologies); Site_1: Not I; Site_2: Sal I; This cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of greenhouse grown plants. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restriction

site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells (Gibco BRL). This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT 66 a 175 c 104 g 129 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 474;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 398 GTTGCTGGATCTGCTG 413

RESULT 25
AL379973
LOCUS AL379973 481 bp mRNA EST 03-AUG-2000
DEFINITION MtBB48E08F1 MtBB Medicago truncatula cDNA clone MtBB48E08 T3, mRNA sequence.
ACCESSION AL379973
VERSION AL379973.1 GI:9679725
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 481)
AUTHORS Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V. and Gamas,P.
TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules
JOURNAL Unpublished (2000)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES Location/Qualifiers
source 1. .481
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MtBB48E08"
/clone_lib="MtBB"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
/note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."
BASE COUNT 89 a 152 c 64 g 176 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 481;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 385 GTTGCTGGATCTGCTG 400

RESULT 26
BF461669/c
LOCUS BF461669 497 bp mRNA EST 04-DEC-2000
DEFINITION UI-M-CG0p-bmu-g-12-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone UI-M-CG0p-bmu-g-12-0-UI 3', mRNA sequence.
ACCESSION BF461669
VERSION BF461669.1 GI:11530852
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 497)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=No.
FEATURES Location/Qualifiers
source 1. .497
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-CG0p-bmu-g-12-0-UI"
/clone_lib="NIH_BMAP_Ret4_S2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_Ret4_S2 library is a subtracted library, ultimately derived from mouse retina tissue libraries at various stages of development. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu.
TAG_SEQ=None found"
BASE COUNT 112 a 168 c 171 g 44 t 2 others
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 497;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16

```
          ||||||||||||||||
Db    246 GTTGCTGGATCTGCTG 231


RESULT  27
BE515828
LOCUS       BE515828      502 bp    mRNA            EST       08-AUG-2000
DEFINITION  WHE0606_B07_C14ZA Wheat ABA-treated embryo cDNA library Triticum
            aestivum cDNA clone WHE0606_B07_C14, mRNA sequence.
ACCESSION   BE515828
VERSION     BE515828.1  GI:9739842
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Anderson,O.D., Chao,S., Han,P.S., Hsia,C.C., Johnson,R.R, Kang,Y.,
            Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Tong,J.C., Verhey
            ,S.D. and Walker-Simmons,M.K.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - ABA-treated embryo library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Clontech Matchmaker 3' AD primer.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Triticum aestivum"
                     /cultivar="Brevor (soft, white, winter, common wheat)"
                     /db_xref="taxon:4565"
                     /clone="WHE0606_B07_C14"
                     /clone_lib="Wheat ABA-treated embryo cDNA library"
                     /tissue_type="Seed embryo"
                     /dev_stage="Mature dormant seeds"
                     /lab_host="E. coli DH12S"
                     /note="Vector: pGAD10; Site_1: EcoRI; Site_2: XhoI;
                     Embryos were cut from mature, dormant seeds and imbibed in
                     25 microM ABA  (abscisic acid) in 5 mM Mes buffer, pH 5.7,
                     for 12 hr at 22 C.  The  tissue, total RNA, and poly(A)
                     RNA were prepared by Steven Verhey in  M.K.
                     Walker-Simmons's lab (USDA-ARS, Washington State Univ.,
                     Pullman,  Washington 99164-6420.  A cDNA library was made
                     by Clontech using a  combination of random and oligo dT
                     primers. Library was plated and  archived by Russell
                     Johnson (Colby College, ME/Walker-Simmons' lab).   Plasmid
                     DNA preparations and DNA sequencing were performed in the
                     OD  Anderson lab (all other authors)."
BASE COUNT      103 a    108 c    128 g    162 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 502;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 atctgctgtttgaagc 24
           ||||||||||||||||
Db     189 ATCTGCTGTTTGAAGC 204


RESULT  28
AL383090
LOCUS       AL383090      503 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBC11G02R1 MtBC Medicago truncatula cDNA clone MtBC11G02 T7, mRNA
            sequence.
ACCESSION   AL383090
VERSION     AL383090.1  GI:9682841
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from endomycorrhizal roots
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBC11G02"
                     /clone_lib="MtBC"
                     /tissue_type="arbuscular mycorrhiza"
                     /dev_stage="harvested 3 weeks post inoculation with Glomus
                     intraradices"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; M. truncatula sterilised seeds were germinated for
                     72h at 25 C, before transplanting into a 1/3 Epoisses soil
                     : 2/3 calcined Terragreen mix in the presence of onion
                     root fragments colonized by the arbuscular mycorrhizal
                     fungus Glomus intraradices (Schenck & Smith, isolate LPA8
                     ). The plants were watered every day and twice a week with
                     a modified nutrient Long Ashton solution without phosphate
                     but with a high level of nitrate. After 3 weeks RNA was
                     extracted from whole root systems. cDNA was prepared from
                     polyA+ enriched RNA. The cDNA was directionally ligated
                     into Uni-zap XR vector from Stratagene and packaged using
                     Gigapack Gold packaging extracts. Plasmids containing cDNA
                     inserts were mass-excised from phage stocks using ExAssit
                     helper phage and propagated in SOLR cells. Clone ordering
                     and sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France). Note : EST may be of
                     fungal origin."
BASE COUNT      107 a    139 c     76 g    181 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 503;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db     237 GTTGCTGGATCTGCTG 252


RESULT  29
BI418981
LOCUS       BI418981      513 bp    mRNA            EST       15-AUG-2001
DEFINITION  LjNEST51h6r Lotus japonicus nodule library 5 and 7 week-old Lotus
            japonicus cDNA 5', mRNA sequence.
ACCESSION   BI418981
VERSION     BI418981.1  GI:15190004
```

```
KEYWORDS    EST.
SOURCE      Lotus japonicus.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
            Lotus.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Colebatch,G., Freund,S., Trevaskis,B and Udvardi,M.
  TITLE     Lotus japonicus root nodule ESTs: tools for functional genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Udvardi MK
            Molecular Plant Nutrition
            Max Planck Institute of Molecular Plant Physiology
            Am Muehlenberg 1, 14476 Golm, Germany
            Fax: 49 331 567 8250
            Email: udvardi@mpimp-golm.mpg.de
            Seq primer: T7
            High quality sequence stop: 513.
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Lotus japonicus"
                     /cultivar="Gifu (B-129)"
                     /db_xref="taxon:34305"
                     /clone_lib="Lotus japonicus nodule library 5 and 7
                     week-old"
                     /dev_stage="5 and 7 week-old plants"
                     /note="Organ: Nodule; Vector: pSPORT1; Site_1: SalI;
                     Site_2: NotI; The library was prepared using mRNA
                     extracted from nodules of 5 and 7 week-old Lotus plants.
                     Nodules were induced by, and contained Mesorhizobium
                     strain R7A."
BASE COUNT       85 a    164 c     99 g    165 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 513;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db     392 GTTGCTGGATCTGCTG 407


RESULT  30
AL373872
LOCUS       AL373872      517 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBB03D01R1 MtBB Medicago truncatula cDNA clone MtBB03D01 T7, mRNA
            sequence.
ACCESSION   AL373872
VERSION     AL373872.1  GI:9673624
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
            nodules
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBB03D01"
                     /clone_lib="MtBB"
                     /tissue_type="symbiotic root nodules"
                     /dev_stage="harvested 4 days post inoculation with
                     Sinorhizobium meliloti"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber on
                     nitrogen-rich medium for 21 days. Three days before
                     inoculation with Sinorhizobium meliloti, the medium was
                     replaced by N-free medium. Root nodules (+ short adjacent
                     root segments) were harvested 4 days post inoculation.
                     cDNA was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zap XR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
BASE COUNT      104 a    147 c     71 g    195 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db     274 GTTGCTGGATCTGCTG 289


RESULT  31
AQ692054
LOCUS       AQ692054      517 bp    DNA             GSS       06-JUL-1999
DEFINITION  HS_5408_A2_E07_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=984 Col=14 Row=I, DNA sequence.
ACCESSION   AQ692054
VERSION     AQ692054.1  GI:5382302
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 984  row: I  column: 14
```

```
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 517.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=984 Col=14 Row=I"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      154 a     96 c     83 g    178 t      6 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 atctgctgtttgaagc 24
           ||||||||||||||||
Db     411 ATCTGCTGTTTGAAGC 426


RESULT  32
BF419551/c
LOCUS       BF419551      521 bp    mRNA            EST       28-NOV-2000
DEFINITION  UI-R-CA0-bpd-c-11-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone
            UI-R-CA0-bpd-c-11-0-UI 3', mRNA sequence.
ACCESSION   BF419551
VERSION     BF419551.1  GI:11407540
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CA0-bpd-c-11-0-UI"
                     /clone_lib="UI-R-CA0"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
                     library is a subtracted library derived from the following
                     tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                     , midbrain,  cerebral cortex, corpus striatum, testis, and
                     hippocampus.  For a detailed description of the library
                     from which this clone was  derived, please visit our web
                     site at ratest.eng.uiowa.edu.  The subtraction has been
                     previously described in (Bonaldo,  Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT      113 a    179 c    183 g     46 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 521;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db     246 GTTGCTGGATCTGCTG 231


RESULT  33
AI854720/c
LOCUS       AI854720      522 bp    mRNA            EST       15-JUL-1999
DEFINITION  UI-M-BH0-akc-f-12-0-UI.s1 NIH_BMAP_M_S1 Mus musculus cDNA clone
            UI-M-BH0-akc-f-12-0-UI 3', mRNA sequence.
ACCESSION   AI854720
VERSION     AI854720.1  GI:5498626
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 522)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .522
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH0-akc-f-12-0-UI"
                     /clone_lib="NIH_BMAP_M_S1"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S1 library is a subtracted library derived from
                     a mixture of normalized libraries from ten regions of the
                     mouse brain (cerebellum, brain stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal ganglia, pineal
                     gland, striatum, hipoccampus). The driver used for
                     subtraction consisted of a pool of 20,000 cDNA clones
                     obtained from non-normalized and normalized libraries of
                     these ten regions of the mouse brain.
                     TAG_LIB=NIH_BMAP_M_S1
                     TAG_TISSUE=hypothalamus
```

```
                    TAG_SEQ=CGGTA"
BASE COUNT      118 a    174 c    179 g     51 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 522;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db    245 GTTGCTGGATCTGCTG 230


RESULT  34
AL379974
LOCUS       AL379974      525 bp    mRNA            EST       03-AUG-2000
DEFINITION  MtBB48E08R1 MtBB Medicago truncatula cDNA clone MtBB48E08 T7, mRNA
            sequence.
ACCESSION   AL379974
VERSION     AL379974.1  GI:9679726
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
            Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
            ,V. and Gamas,P.
  TITLE     Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
            nodules
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
            Biologie Moleculaire des Relations Plantes-Microorganismes,
            CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
            Mt-est@toulouse.inra.fr Website :
            http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES             Location/Qualifiers
     source          1. .525
                     /organism="Medicago truncatula"
                     /cultivar="Jemalong"
                     /db_xref="taxon:3880"
                     /clone="MtBB48E08"
                     /clone_lib="MtBB"
                     /tissue_type="symbiotic root nodules"
                     /dev_stage="harvested 4 days post inoculation with
                     Sinorhizobium meliloti"
                     /note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
                     XhoI; Plants were grown in an aeroponic chamber on
                     nitrogen-rich medium for 21 days. Three days before
                     inoculation with Sinorhizobium meliloti, the medium was
                     replaced by N-free medium. Root nodules (+ short adjacent
                     root segments) were harvested 4 days post inoculation.
                     cDNA was prepared from polyA+ enriched RNA. The cDNA was
                     directionally ligated into Uni-zap XR vector from
                     Stratagene and packaged using Gigapack Gold packaging
                     extracts. Plasmids containing cDNA inserts were
                     mass-excised from phage stocks using ExAssit helper phage
                     and propagated in SOLR cells. Clone ordering and
                     sequencing was performed by the Centre National de
                     Sequencage (Genoscope, Evry, France)."
BASE COUNT      108 a    148 c     81 g    188 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 525;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db    255 GTTGCTGGATCTGCTG 270


RESULT  35
AW395303
LOCUS       AW395303      531 bp    mRNA            EST       17-JUL-2000
DEFINITION  sh46d09.y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1017-4746 5' similar to SW:RPB9_DROME P36958 DNA-DIRECTED RNA
            POLYMERASE II 15.1 KD POLYPEPTIDE ;, mRNA sequence.
ACCESSION   AW395303
VERSION     AW395303.1  GI:6913773
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Insert Length: 743   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 342.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1017-4746"
                     /clone_lib="Gm-c1017"
                     /tissue_type="vegetable buds of field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from vegatable buds of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT      159 a     96 c    107 g    169 t
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 10;  Length 531;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db      60 GTTGCTGGATCTGCTG 75


RESULT  36
AQ889100/c
LOCUS       AQ889100      555 bp    DNA             GSS       10-NOV-1999
DEFINITION  HS_2204_A1_C07_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2204 Col=13 Row=E, DNA sequence.
ACCESSION   AQ889100
VERSION     AQ889100.1  GI:6345290
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2204  row: E  column: 13
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 555.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2204 Col=13 Row=E"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      131 a    135 c    147 g    124 t     18 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 555;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 tctgctgtttgaagcg 25
           ||||||||||||||||
Db     497 TCTGCTGTTTGAAGCG 482


RESULT  37
BI229421/c
LOCUS       BI229421      570 bp    mRNA            EST       11-JUL-2001
DEFINITION  RE27194.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE27194 5 similar to CG1148:
            FBan0001148 located on: 3R 83E1-83E1;: 04/12/2001, mRNA sequence.
ACCESSION   BI229421
VERSION     BI229421.1  GI:14696685
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson
            ,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George
            ,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,
            Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin
            ,G.M.
  TITLE     BDGP/HHMI RE Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003600: arm:3R [1868408,2159952]
            estimated-cyto:83D2-83E4: 04/12/2001
            Plate: RE.271  row: H  column: 10
            High quality sequence stop: 545.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RE27194"
                     /clone_lib="RE Drosophila melanogaster normalized Embryo
                     pFlc-1"
                     /sex="male and female"
                     /dev_stage="0-24 hours mixed stage embryonic"
                     /lab_host="DH5-alpha TonA"
                     /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
                     BamHI; Library was kindly generated by Piero Carninci at
                     the RIKEN. The library  was normalized and excised using
                     Cre recombinase. Plasmid cDNA library."
BASE COUNT      139 a    179 c    153 g     99 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 570;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db      63 GTTGCTGGATCTGCTG 48


RESULT  38
AW573807
LOCUS       AW573807      574 bp    mRNA            EST       07-SEP-2000
DEFINITION  EST316398 GVN Medicago truncatula cDNA clone pGVN-48O22, mRNA
            sequence.
ACCESSION   AW573807
VERSION     AW573807.1  GI:7238540
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng
            ,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
            Holt,I.E. and Fraser,C.M.
  TITLE     ESTs from one month old nitrogen-fixing root nodules of Medicago
            truncatula
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Carroll P. Vance
```

```
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            Minnesota EST name:M252705e
            TIGR sequence name:MTCAF95TK
            More information is available at. .
            'http://chrysie.tamu.edu/medicago'
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES              Location/Qualifiers
     source           1. .574
                      /organism="Medicago truncatula"
                      /cultivar="genotype A17"
                      /db_xref="taxon:3880"
                      /clone="pGVN-48O22"
                      /clone_lib="GVN"
                      /tissue_type="N2-fixing root nodules"
                      /dev_stage="effective root nodules harvested one month
                      post inoculation with Sinorhizobium meliloti"
                      /lab_host="E. coli strain XLOLR"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; cDNA was prepared from polyA+ enriched RNA from
                      effective root nodules harvested one month post
                      inoculation with Sinorhizobium meliloti. The cDNA was
                      directionally ligated into the Uni-ZAP XR vector from
                      Stratagene and packaged using Gigapack III Gold packaging
                      extracts. Plasmids containing cDNA inserts were excised
                      from the recombinant lambda-ZAP phage using Ex-Assist
                      helper phage and propagated in XLOLR cells."
BASE COUNT      110 a    165 c     80 g    218 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 574;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db    370 GTTGCTGGATCTGCTG 385


RESULT  39
BE999070
LOCUS       BE999070      575 bp    mRNA            EST       06-OCT-2000
DEFINITION  EST430793 GVSN Medicago truncatula cDNA clone pGVSN-15C19, mRNA
            sequence.
ACCESSION   BE999070
VERSION     BE999070.1  GI:10699346
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town
            ,C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
  TITLE     ESTs from senescent nodules of Medicago truncatula
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            University of Minnesota name: M273048e TIGR sequence name:
            MTKAY22TK More information is available at:
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES              Location/Qualifiers
     source           1. .575
                      /organism="Medicago truncatula"
                      /cultivar="genotype A17"
                      /db_xref="taxon:3880"
                      /clone="pGVSN-15C19"
                      /clone_lib="GVSN"
                      /tissue_type="senescent root nodules"
                      /dev_stage="mixture of effective nodules from 40 day old
                      plants harvested 36 hours post shoot removal and nodules
                      collected from 2 month old plants at mid-pod stage"
                      /lab_host="E. coli strain SOLR"
                      /note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
                      XhoI; cDNA was prepared from polyA+ enriched RNA from the
                      mixture of effective nodules of 40 day old plants
                      harvested 36 hours post shoot removal and nodules
                      collected from 2 month old plants at mid-pod stage. The
                      cDNA was directionally ligated into the Uni-ZAP XR vector
                      from Stratagene and packaged using Gigapack III Gold
                      packaging extracts. Plasmids containing cDNA inserts were
                      excised from the recombinant lambda-ZAP phage using
                      Ex-Assist helper phage and propagated in SOLR cells."
BASE COUNT      103 a    164 c     82 g    226 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 575;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gttgctggatctgctg 16
          ||||||||||||||||
Db    347 GTTGCTGGATCTGCTG 362


RESULT  40
BF525150
LOCUS       BF525150      580 bp    mRNA            EST       11-DEC-2000
DEFINITION  UI-R-AF0-yd-e-05-0-UI.r1 UI-R-AF0 Rattus norvegicus cDNA clone
            UI-R-AF0-yd-e-05-0-UI 5', mRNA sequence.
ACCESSION   BF525150
VERSION     BF525150.1  GI:11633117
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1795307
            Seq primer: M13 Forward.
FEATURES              Location/Qualifiers
     source           1. .580
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-AF0-yd-e-05-0-UI"
```

/clone_lib="UI-R-AF0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AF0 library is a non-normalized library constructed from 15 dpc rat atrioventricular (AV) canal. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806 , 1996. Tissue provided by Jim Lin, Department of Biology, University of Iowa."
BASE COUNT 155 a 146 c 147 g 132 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 580;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tgctggatctgctgtt 18
||||||||||||||||
Db 321 TGCTGGATCTGCTGTT 336

RESULT 41
AW734894
LOCUS AW734894 591 bp mRNA EST 24-APR-2000
DEFINITION sk76d03.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1016-10014 5', mRNA sequence.
ACCESSION AW734894
VERSION AW734894.1 GI:7640531
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
REFERENCE 1 (bases 1 to 591)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 380.
FEATURES Location/Qualifiers
source 1. .591
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1016-10014"
/clone_lib="Gm-c1016"
/tissue_type="immature flowers of field grown plants"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; This cDNA library was constructed from mRNA isolated from immature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelding."
BASE COUNT 121 a 180 c 103 g 185 t 2 others
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 591;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 330 GTTGCTGGATCTGCTG 345

RESULT 42
AZ267971
LOCUS AZ267971 601 bp DNA GSS 26-JUL-2000
DEFINITION RPCI-23-103L11.TJ RPCI-23 Mus musculus genomic clone RPCI-23-103L11 , DNA sequence.
ACCESSION AZ267971
VERSION AZ267971.1 GI:9481584
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 601)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-103L11.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 103 row: L column: 11
Seq primer: SP6
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .601
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-103L11"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the

EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "
BASE COUNT 189 a 92 c 111 g 209 t
ORIGIN

Query Match 64.0%; Score 16; DB 13; Length 601;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ttgctggatctgctgt 17
||||||||||||||||
Db 354 TTGCTGGATCTGCTGT 369

RESULT 43
BI229406/c
LOCUS BI229406 602 bp mRNA EST 11-JUL-2001
DEFINITION RE27178.5prime RE Drosophila melanogaster normalized Embryo pFlc-1 Drosophila melanogaster cDNA clone RE27178 5 similar to CG1148: FBan0001148 located on: 3R 83E1-83E1;: 04/12/2001, mRNA sequence.
ACCESSION BI229406
VERSION BI229406.1 GI:14696670
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 602)
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.
TITLE BDGP/HHMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic AE003600: arm:3R [1868408,2159952]
estimated-cyto:83D2-83E4: 04/12/2001
Plate: RE.271 row: G column: 6
High quality sequence stop: 476.
FEATURES Location/Qualifiers
source 1. .602
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="RE27178"
/clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha TonA"
/note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
BASE COUNT 148 a 185 c 161 g 108 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 602;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 63 GTTGCTGGATCTGCTG 48

RESULT 44
AW267738
LOCUS AW267738 609 bp mRNA EST 07-SEP-2000
DEFINITION EST305866 DSIR Medicago truncatula cDNA clone pDSIR-7I7, mRNA sequence.
ACCESSION AW267738
VERSION AW267738.1 GI:6654694
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 609)
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,G.S., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula after inoculation with Phytophthora medicaginis
JOURNAL Unpublished (1999)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
Minnesota EST name:M250222e
TIGR sequence name:MTBAA52TK
More information, including clone ordering, is available at. .
http://chrysie.tamu.edu/medicago
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES Location/Qualifiers
source 1. .609
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pDSIR-7I7"
/clone_lib="DSIR"
/tissue_type="roots infected with Phytophthora medicaginis"
/dev_stage="roots harvested at 10 days post inoculation with Phytophthora medicaginis"
/lab_host="E. coli strain XLOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10 days post inoculation with Phytophthora medicaginis. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells. Note: EST may be of fungal origin."
BASE COUNT 117 a 172 c 82 g 238 t
ORIGIN

Query Match 64.0%; Score 16; DB 10; Length 609;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gttgctggatctgctg 16
||||||||||||||||
Db 381 GTTGCTGGATCTGCTG 396

RESULT 45
AW980891
LOCUS AW980891 616 bp mRNA EST 07-SEP-2000

```
DEFINITION  EST392044 GVN Medicago truncatula cDNA clone pGVN-60C11, mRNA
            sequence.
ACCESSION   AW980891
VERSION     AW980891.1  GI:8172436
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng
            ,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
            Holt,I.E. and Fraser,C.M.
  TITLE     ESTs from one month old nitrogen-fixing root nodules of Medicago
            truncatula
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            University of Minnesota name:M256725e
            TIGR sequence name:MTCBW18TK
            More information is available at. . http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .616
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pGVN-60C11"
                     /clone_lib="GVN"
                     /tissue_type="N2-fixing root nodules"
                     /dev_stage="effective root nodules harvested one month
                     post inoculation with Sinorhizobium meliloti"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from
                     effective root nodules harvested one month post
                     inoculation with Sinorhizobium meliloti. The cDNA was
                     directionally ligated into the Uni-ZAP XR vector from
                     Stratagene and packaged using Gigapack III Gold packaging
                     extracts. Plasmids containing cDNA inserts were excised
                     from the recombinant lambda-ZAP phage using Ex-Assist
                     helper phage and propagated in XLOLR cells."
BASE COUNT      118 a    173 c     82 g    243 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 616;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gttgctggatctgctg 16
           ||||||||||||||||
Db     388 GTTGCTGGATCTGCTG 403



Search completed: December 26, 2001, 14:10:04
Job time: 7051 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:50:57 ; Search time 1757.07 Seconds
(without alignments)
234.726 Million cell updates/sec

Title: US-09-396-196F-4
Perfect score: 25
Sequence: 1 ggatctgctgtttgaagcgcagcag 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 9

Total number of hits satisfying chosen parameters: 252155

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		25	100.0	1041	6	AR029499	AR029499 Sequence
2		25	100.0	1041	6	AR034916	AR034916 Sequence
3		25	100.0	1084	6	A11530	A11530 BioB gene o
4		25	100.0	5526	1	AF250776	AF250776 Unculture
5		25	100.0	5793	1	ECOBIO	J04423 E.coli 7,8-
6		25	100.0	5872	6	A38246	A38246 Sequence 1
7		25	100.0	5872	6	A38251	A38251 Sequence 6
8		25	100.0	5872	6	A93674	A93674 Sequence 1
9		25	100.0	5872	6	A93679	A93679 Sequence 6
10		25	100.0	5872	6	AR101809	AR101809 Sequence
11		25	100.0	5872	6	AR101810	AR101810 Sequence
12		25	100.0	11022	1	AE000180	AE000180 Escherich
13		25	100.0	12891	1	AE004192	AE004192 Vibrio ch
14		25	100.0	13501	1	AE005258	AE005258 Escherich
15		25	100.0	297816	1	AP002553	AP002553 Escherich
16		21	84.0	128	1	ECOBIOB	M27731 Escherichia
17		21	84.0	8227	1	AF248314	AF248314 Unculture
18		18	72.0	1041	1	EHBIOTOP2	U38648 Erwinia her
19		18	72.0	1121	6	E00893	E00893 Genomic DNA
20	c	17	68.0	1044	5	AF250347	AF250347 Xenopus l
21	c	17	68.0	1614	5	AF281080	AF281080 Xenopus l
22		17	68.0	14268	1	AE001220	AE001220 Treponema
23	c	16	64.0	801	6	AX120335	AX120335 Sequence
24	c	16	64.0	924	6	AX064067	AX064067 Sequence
25		16	64.0	6887	8	YLI012084	AJ012084 Yarrowia
26		16	64.0	11823	1	AE002054	AE002054 Deinococc
27		16	64.0	13856	1	AE004026	AE004026 Xylella f
28	c	16	64.0	40031	3	CBRG01D9	U56248 Caenorhabdi
29	c	16	64.0	156425	8	AP002897	AP002897 Oryza sat
30	c	16	64.0	163489	9	AC016697	AC016697 Homo sapi
31	c	16	64.0	168177	2	AC023914	AC023914 Homo sapi
32		16	64.0	171069	2	AP003845	AP003845 Oryza sat
33		16	64.0	191677	2	AC090534	AC090534 Mus muscu
34	c	16	64.0	200000	2	AC007104	AC007104 Homo sapi
35	c	16	64.0	349980	6	AX120085	AX120085 Sequence
36		15	60.0	306	14	PPHE7D	D50548 Human papil
37		15	60.0	625	8	PEU244276	AJ244276 Pringshei
38		15	60.0	810	6	AX110623	AX110623 Sequence
39		15	60.0	1551	1	AF128956	AF128956 Escherich
40		15	60.0	1572	1	SMAAAC3VB	M97172 Serratia ma
41	c	15	60.0	1616	10	AF239157	AF239157 Rattus no
42	c	15	60.0	1623	10	AF009246	AF009246 Mus muscu
43		15	60.0	2795	1	AF039572	AF039572 Serratia
44	c	15	60.0	6043	14	PMO238607	AJ238607 Potato mo
45		15	60.0	7779	14	HPU31779	U31779 Human papil
46	c	15	60.0	8219	7	AF066865	AF066865 Bacteriop
47		15	60.0	10881	1	AE004129	AE004129 Vibrio ch
48		15	60.0	11062	1	AE006532	AE006532 Streptoco
49	c	15	60.0	11180	1	AE007575	AE007575 Clostridi
50	c	15	60.0	11328	1	AE004322	AE004322 Vibrio ch
51	c	15	60.0	11467	1	AE006281	AE006281 Lactococc
52	c	15	60.0	11674	1	AE005391	AE005391 Escherich
53	c	15	60.0	11684	1	AE000263	AE000263 Escherich
54		15	60.0	13823	1	D90812	D90812 E.coli geno
55		15	60.0	19521	1	D90811	D90811 E.coli geno
56	c	15	60.0	36949	7	AF323670	AF323670 Bacteriop
57		15	60.0	37555	3	U42832	U42832 Caenorhabdi
58	c	15	60.0	37991	9	AC005101	AC005101 Homo sapi
59	c	15	60.0	38245	1	AF105019	AF105019 Salmonell
60		15	60.0	40629	2	AC006105	AC006105 Homo sapi
61	c	15	60.0	42563	8	AB018110	AB018110 Arabidops
62	c	15	60.0	43439	2	U82208	U82208 Homo sapien
63		15	60.0	43760	2	AC087367	AC087367 Homo sapi
64	c	15	60.0	43955	2	AC015342	AC015342 Drosophil
65	c	15	60.0	46387	2	U82212	U82212 Homo sapien
66		15	60.0	60248	9	AL137181	AL137181 Human DNA
67		15	60.0	60310	9	AL391647	AL391647 Human DNA
68		15	60.0	62544	8	NCG65A3	AL513411 Neurospor
69	c	15	60.0	63633	9	AL139409	AL139409 Human DNA
70		15	60.0	69918	9	AL441988	AL441988 Human DNA

c	71	15	60.0	84107	9	HS248E1
	72	15	60.0	85084	9	AB042297
c	73	15	60.0	86516	9	AC009493
	74	15	60.0	90720	2	AC025390
	75	15	60.0	91201	9	HSDA86H19
c	76	15	60.0	106408	9	AL512303
c	77	15	60.0	110000	2	LMFLCHR32_21
	78	15	60.0	110000	2	LMFLCHR36_30
	79	15	60.0	110000	2	AC020850_1
	80	15	60.0	110716	9	AL135918
	81	15	60.0	125287	9	HS159A19
c	82	15	60.0	128589	2	AC087122
	83	15	60.0	129048	2	AL592286
c	84	15	60.0	132592	2	AL589792
	85	15	60.0	134137	9	HS3E5
	86	15	60.0	136930	2	AL590151
c	87	15	60.0	137733	2	AC026292
	88	15	60.0	138097	2	AC027787
c	89	15	60.0	138097	2	AC027787
c	90	15	60.0	146349	2	AC007863
	91	15	60.0	148891	9	AL162274
c	92	15	60.0	148927	2	AC087127
c	93	15	60.0	150237	2	AL357335
	94	15	60.0	151494	2	AL138931
c	95	15	60.0	152755	2	AC006358
c	96	15	60.0	154005	2	AC073208
	97	15	60.0	157442	2	AC079929
	98	15	60.0	161616	2	AC084824
c	99	15	60.0	161635	9	AC083873
c	100	15	60.0	161651	9	HS233K16

AL023578	Human DNA
AB042297	Homo sapi
AC009493	Homo sapi
AC025390	Homo sapi
AL049562	Human DNA
AL512303	Human DNA
Continuation (22 o	
Continuation (31 o	
Continuation (2 of	
AL135918	Human DNA
AL031729	Human DNA
AC087122	Mus muscu
AL592286	Homo sapi
AL589792	Homo sapi
AL022239	Human DNA
AL590151	Danio rer
AC026292	Homo sapi
AC027787	Homo sapi
AC027787	Homo sapi
AC007863	Trypanoso
AL162274	Human DNA
AC087127	Mus muscu
AL357335	Homo sapi
AL138931	Homo sapi
AC006358	Homo sapi
AC073208	Homo sapi
AC079929	Homo sapi
AC084824	Homo sapi
AC083873	Homo sapi
AL035406	Human DNA

ALIGNMENTS

```
RESULT   1
AR029499
LOCUS       AR029499      1041 bp    DNA              PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1  GI:5941472
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.Andrew.
  TITLE     Enhanced biotin biosynthesis in plant tissue
  JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.00012;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagcgcagcag 25
          |||||||||||||||||||||||||
Db     63 GGATCTGCTGTTTGAAGCGCAGCAG 87


RESULT   2
AR034916
LOCUS       AR034916      1041 bp    DNA              PAT       29-SEP-1999
DEFINITION  Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1  GI:5950521
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Patton,D.A.
  TITLE     Transgenic plants having increased biotin content
  JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="unknown"
BASE COUNT      262 a    273 c    305 g    201 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 0.00012;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagcgcagcag 25
          |||||||||||||||||||||||||
Db     63 GGATCTGCTGTTTGAAGCGCAGCAG 87


RESULT   3
A11530
LOCUS       A11530        1084 bp    DNA              PAT       09-FEB-1994
DEFINITION  BioB gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1  GI:490218
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1084)
  AUTHORS   .
  JOURNAL   Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES             Location/Qualifiers
     source          1. .1084
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            24. .1064
                     /gene="BIOB"
     CDS             24. .1064
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA00965.1"
                     /db_xref="GI:490219"
                     /db_xref="SWISS-PROT:P12996"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
BASE COUNT      271 a    286 c    318 g    209 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0.00012;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagcgcagcag 25
          |||||||||||||||||||||||||
Db     86 GGATCTGCTGTTTGAAGCGCAGCAG 110


RESULT   4
AF250776
LOCUS       AF250776      5526 bp    DNA              BCT       31-JAN-2001
DEFINITION  Uncultured bacterium pCosHE2 hypothetical 17.1 kDa protein in
```

```
          modC-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
          biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
          biotin biosynthesis protein BioC (bioC) genes, complete cds; and
          dethiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION AF250776
VERSION   AF250776.1  GI:12620124
KEYWORDS  .
SOURCE    uncultured bacterium pCosHE2.
  ORGANISM uncultured bacterium pCosHE2
          Bacteria; environmental samples.
REFERENCE 1  (bases 1 to 5526)
  AUTHORS Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE   Direct cloning from enrichment cultures, a reliable strategy for
          isolation of complete operons and genes from microbial consortia
  JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
  MEDLINE 20575196
   PUBMED 11133432
REFERENCE 2  (bases 1 to 5526)
  AUTHORS Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE   Direct Submission
  JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
          Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .5526
                     /organism="uncultured bacterium pCosHE2"
                     /db_xref="taxon:143797"
                     /clone="pCosHE2"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     CDS             complement(52. .528)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical 17.1 kDa protein in modC-bioA
                     intergenic region"
                     /protein_id="AAG60577.1"
                     /db_xref="GI:12620125"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKAG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(587. .1876)
                     /gene="bioA"
     CDS             complement(587. .1876)
                     /gene="bioA"
                     /note="7,8-diaminopelargonic acid
                     synthetase-aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA-aminotransferase BioA"
                     /protein_id="AAG60578.1"
                     /db_xref="GI:12620126"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LKCVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            1963. .3003
                     /gene="bioB"
     CDS             1963. .3003
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG60579.1"
                     /db_xref="GI:12620127"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPNERDMPYLEQMVQGVKDLGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3000. .4154
                     /gene="bioF"
     CDS             3000. .4154
                     /gene="bioF"
                     /note="8-amino-7-oxononanoate synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="KAPA synthetase BioF"
                     /protein_id="AAG60580.1"
                     /db_xref="GI:12620128"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQLVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRRDEGDARREKLVSLIAHFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4141. .4896
                     /gene="bioC"
     CDS             4141. .4896
                     /gene="bioC"
                     /note="reaction step prior to synthesis of pimeloyl-CoA"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis protein BioC"
                     /protein_id="AAG60581.1"
                     /db_xref="GI:12620129"
                     /translation="MATVNKQAIAAAFGRAAALYEQHADLQRQSADVLLAMLPERKYT
                     HVLDAGCGPGWMTRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRSKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4883. .>5526
                     /gene="bioD"
     CDS             4883. .>5526
                     /gene="bioD"
                     /note="DTB synthetase"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase BioD"
                     /protein_id="AAG60582.1"
                     /db_xref="GI:12620130"
                     /translation="MLVSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVAS
                     GSEKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESSVMS
                     SGLRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAML
                     TAQAIQHAGLTLAGWVANDVTPPGKRHAEYITTLTRMIPAPLLGEIPWLAENPE"
BASE COUNT     1274 a   1507 c   1567 g   1178 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 1;  Length 5526;
  Best Local Similarity   100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db    2025 GGATCTGCTGTTTGAAGCGCAGCAG 2049


RESULT   5
ECOBIO
LOCUS       ECOBIO       5793 bp    DNA             BCT       28-FEB-1994
DEFINITION  E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
            (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioC
            protein, and dethiobiotin synthetase (bioD), complete cds.
ACCESSION   J04423
VERSION     J04423.1  GI:145422
KEYWORDS    7,8-diamino-pelargonic acid aminotransferase;
            7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
            bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
            synthetase.
```

```
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5793)
  AUTHORS   Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
            Johnson,O.
  TITLE     The Eschierichia coli biotin biosynthetic enzyme sequences
            predicted
  JOURNAL   J. Biol. Chem. 263, 19577-19585 (1988)
  MEDLINE   89066784
COMMENT     Draft entry and computer-readable sequence [1] kindly submitted by
            A.Otsuka, 09-NOV-1988.
FEATURES             Location/Qualifiers
     source          1. .5793
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
     CDS             complement(98. .574)
                     /note="ORF 1"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23513.1"
                     /db_xref="GI:455168"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASAAMVGFNVHFHSLASASITAMFS"
     gene            complement(633. .1925)
                     /gene="bioA"
     CDS             complement(633. .1925)
                     /gene="bioA"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diamino-pelargonic acid aminotransferase"
                     /protein_id="AAA23514.1"
                     /db_xref="GI:457106"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMSGRN
                     ALECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDN
                     SMHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGA
                     GGMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKA
                     LTGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQ
                     QQVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIR
                     PFGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     gene            2012. .3052
                     /gene="bioB"
     CDS             2012. .3052
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAA23515.1"
                     /db_xref="GI:145425"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQTSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     gene            3049. .4203
                     /gene="bioF"
     CDS             3049. .4203
                     /gene="bioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-keto-8-amino pelargonic acid synthetase"
                     /protein_id="AAA23516.1"
                     /db_xref="GI:145426"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     gene            4190. .4945
                     /gene="bioC"
     CDS             4190. .4945
                     /gene="bioC"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA23517.1"
                     /db_xref="GI:145427"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPERHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     gene            4938. .5597
                     /gene="bioD"
     CDS             4938. .5597
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAA23518.1"
                     /db_xref="GI:145428"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKRAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPRRCWERSPGLQKIQKMRQPES
                     T"
BASE COUNT     1363 a   1554 c   1631 g   1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.


  Query Match           100.0%;  Score 25;  DB 1;  Length 5793;
  Best Local Similarity  100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db    2074 GGATCTGCTGTTTGAAGCGCAGCAG 2098


RESULT   6
A38246
LOCUS       A38246        5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 1 from Patent WO9408023.
ACCESSION   A38246
VERSION     A38246.1  GI:2294844
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 1 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
```

```
promoter        1. .96
                /function="PROMOTER PTAC"
                /evidence=experimental
-35_signal      23. .28
-10_signal      45. .49
                /standard_name="PROMOTER PTAC"
                /evidence=experimental
RBS             105. .119
                /standard_name="BIOB RBS NO.9"
                /evidence=experimental
gene            117. .1157
                /gene="BIOB"
CDS             117. .1157
                /gene="BIOB"
                /codon_start=1
                /transl_table=11
                /number=1
                /evidence=experimental
                /product="BIOTIN SYNTHASE"
                /protein_id="CAA02325.1"
                /db_xref="GI:2294845"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
RBS             2284. .2297
                /standard_name="BIOC RBS"
gene            2295. .3050
                /gene="BIOC"
CDS             2295. .3050
                /gene="BIOC"
                /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                /codon_start=1
                /transl_table=11
                /number=3
                /product="PROTEIN"
                /protein_id="CAA02326.1"
                /db_xref="GI:2294846"
                /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
RBS             3742. .3752
                /standard_name="BIOA RBS"
gene            3750. .5039
                /gene="BIOA"
CDS             3750. .5039
                /gene="BIOA"
                /standard_name="S-ADENOSYL-L-METHIONINE:
                8-AMINO-7-OXONONANOATE AMINOTRANSF."
                /EC_number="2.6.1.62"
                /codon_start=1
                /transl_table=11
                /number=5
                /evidence=experimental
                /product="DAPA SYNTHASE"
                /protein_id="CAA02327.1"
                /db_xref="GI:2294847"
                /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
RBS             5088. .5100
                /standard_name="ORFI RBS"
gene            5098. .5574
                /gene="ORFI"
CDS             5098. .5574
                /gene="ORFI"
                /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                /codon_start=1
                /transl_table=11
                /number=6
                /evidence=experimental
                /product="PROTEIN"
                /protein_id="CAA02328.1"
                /db_xref="GI:2294848"
                /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
stem_loop       5583. .5605
terminator      5583. .5644
                /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity  100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagcgcagcag 25
          |||||||||||||||||||||||||
Db    179 GGATCTGCTGTTTGAAGCGCAGCAG 203


RESULT   7
A38251
LOCUS       A38251       5872 bp    DNA             PAT       05-MAR-1997
DEFINITION  Sequence 6 from Patent WO9408023.
ACCESSION   A38251
VERSION     A38251.1  GI:2294849
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
  JOURNAL   Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /number=2
                     /evidence=experimental
```

```
                     /product="KAPA SYNTHASE"
                     /protein_id="CAA02329.1"
                     /db_xref="GI:2294850"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /number=4
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAA02330.1"
                     /db_xref="GI:2294851"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match           100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity  100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db     179 GGATCTGCTGTTTGAAGCGCAGCAG 203


RESULT   8
A93674
LOCUS       A93674       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 1 from Patent EP0798384.
ACCESSION   A93674
VERSION     A93674.1  GI:6741862
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A-15/9"
     promoter        1. .96
                     /function="PROMOTER PTAC"
                     /evidence=experimental
     -35_signal      23. .28
     -10_signal      45. .49
                     /standard_name="PROMOTER PTAC"
                     /evidence=experimental
     RBS             105. .119
                     /standard_name="BIOB RBS NO.9"
                     /evidence=experimental
     gene            117. .1157
                     /gene="BIOB"
     CDS             117. .1157
                     /gene="BIOB"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="BIOTIN SYNTHASE"
                     /protein_id="CAB69590.1"
                     /db_xref="GI:6741863"
                     /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                     PDTDEYYNAAAL"
     RBS             2284. .2297
                     /standard_name="BIOC RBS"
     gene            2295. .3050
                     /gene="BIOC"
     CDS             2295. .3050
                     /gene="BIOC"
                     /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /product="PROTEIN"
                     /protein_id="CAB69591.1"
                     /db_xref="GI:6741864"
                     /translation="MATVNKQAIAAAFGRAAAHYEQHADLQRQSADALLAMLPQRKYT
                     HVLDAGCGPGWMSRHWRERHAQVTALDLSPPMLVQARQKDAADHYLAGDIESLPLATA
                     TFDLAWSNLAVQWCGNLSTALRELYRVVRPKGVVAFTTLVQGSLPELHQAWQAVDERP
                     HANRFLPPDEIEQSLNGVHYQHHIQPITLWFDDALSAMRSLKGIGATHLHEGRDPRIL
                     TRSQLQRLQLAWPQQQGRYPLTYHLFLGVIARE"
     RBS             3742. .3752
                     /standard_name="BIOA RBS"
     gene            3750. .5039
                     /gene="BIOA"
     CDS             3750. .5039
                     /gene="BIOA"
                     /standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
                     OXONONANOATE AMINOTRANSF."
                     /EC_number="2.6.1.62"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DAPA SYNTHASE"
                     /protein_id="CAB69592.1"
                     /db_xref="GI:6741865"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
                     QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                     FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     RBS             5088. .5100
                     /standard_name="ORFI RBS"
     gene            5098. .5574
                     /gene="ORFI"
     CDS             5098. .5574
                     /gene="ORFI"
                     /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="PROTEIN"
                     /protein_id="CAB69593.1"
                     /db_xref="GI:6741866"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
```

```
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     stem_loop       5583. .5605
     terminator      5583. .5644
                     /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
                     TERMINATOR"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagcgcagcag 25
          |||||||||||||||||||||||||
Db    179 GGATCTGCTGTTTGAAGCGCAGCAG 203


RESULT   9
A93679
LOCUS       A93679       5872 bp    DNA             PAT       22-JAN-2000
DEFINITION  Sequence 6 from Patent EP0798384.
ACCESSION   A93679
VERSION     A93679.1  GI:6741867
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O. and Brass,J.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="Escherichia coli"
                     /strain="DSM498"
                     /db_xref="taxon:562"
                     /clone="PBO30A15-9"
     RBS             1141. .1156
                     /standard_name="BIOF RBS"
     gene            1154. .2308
                     /gene="BIOF"
     CDS             1154. .2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
                     /EC_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                     SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                     LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                     LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                     SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL
                     RASLAVIRSDEGDARREKLAALITRFRAGVQDLPFTLADSCSAIQPLIVGDNSRALQL
                     AEKLRQQGCWVTAIRPPTVPAGTARLRLTLTAAHEMQDIDRLLEVLHGNG"
     RBS             3030. .3045
                     /standard_name="BIOD RBS"
     gene            3043. .3753
                     /gene="BIOD"
     CDS             3043. .3753
                     /gene="BIOD"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /EC_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAAGYRTAGYKPVASGS
                     EKTPEGLRNSDALALQRNSSLQLDYATVNPYTFAEPTSPHIISAQEGRPIESLVMSAG
                     LRALEQQADWVLVEGAGGWFTPLSDTFTFADWVTQEQLPVILVVGVKLGCINHAMLTA
                     QVIQHAGLTLAGWVANDVTPPGKRHAEYMTTLTRMIPAPLLGEIPWLAENPENAATGK
                     YINLAFVDASTLGFTSRL"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagcgcagcag 25
          |||||||||||||||||||||||||
Db    179 GGATCTGCTGTTTGAAGCGCAGCAG 203


RESULT  10
AR101809
LOCUS       AR101809     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6083712.
ACCESSION   AR101809
VERSION     AR101809.1  GI:12812607
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagcgcagcag 25
          |||||||||||||||||||||||||
Db    179 GGATCTGCTGTTTGAAGCGCAGCAG 203


RESULT  11
AR101810
LOCUS       AR101810     5872 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 6 from patent US 6083712.
ACCESSION   AR101810
VERSION     AR101810.1  GI:12812608
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5872)
  AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
  TITLE     Biotechnological method of producing biotin
  JOURNAL   Patent: US 6083712-A 6 04-JUL-2000;
FEATURES             Location/Qualifiers
     source          1. .5872
                     /organism="unknown"
BASE COUNT     1318 a   1552 c   1695 g   1307 t
ORIGIN


  Query Match             100.0%;  Score 25;  DB 6;  Length 5872;
```

```
  Best Local Similarity   100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcaag 25
           |||||||||||||||||||||||||
Db     179 GGATCTGCTGTTTGAAGCGCAGCAG 203


RESULT  12
AE000180
LOCUS       AE000180    11022 bp    DNA             BCT       01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 70 of 400 of the complete
            genome.
ACCESSION   AE000180 U00096
VERSION     AE000180.1  GI:1786988
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11022)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 11022)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 11022)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .11022
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     repeat_region   <1. .130
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 4 REP sequences"
     gene            complement(147. .1430)
                     /gene="ybhC"
                     /note="b0772"
     CDS             complement(147. .1430)
                     /gene="ybhC"
                     /function="orf; Not classified"
                     /note="f427; 98 pct identical to fragment YBHC_ECOLI
                     SW:P46130 (300 aa) but contains 127 additional C-terminal
                     residues"
                     /codon_start=1
                     /transl_table=11
                     /product="putative pectinesterase"
                     /protein_id="AAC73859.1"
                     /db_xref="GI:1786989"
                     /translation="MNTFSVSRLALALAFGVTLTACSSTPPDQRPSDQTAPGTSSRPI
                     LSAKEAQNFDAQHYFASLTPGAAAWNPSPITLPAQPDFVVGPAGTQGVTHTTIQAAVD
                     AAIIKRTNKRQYIAVMPGEYQGTVYVPAAPGGITLYGTGEKPIDVKIGLSLDGGMSPA
                     DWRHDVNPRGKYMPGKPAWYMYDSCQSKRSDSIGVLCSAVFWSQNNGLQLQNLTIENT
                     LGDSVDAGNHPAVALRTDGDQVQINNVNILGRQNTFFVTNSGVQNRLETNRQPRTLVT
                     NSYIEGDVDIVSGRGAVVFDNTEFRVVNSRTQQEAYVFAPATLSNIYYGFLAVNSRFN
                     AFGDGVAQLGRSLDVDANTNGQVVIRDSAINEGFNTAKPWADAVISNRPFAGNTGSVD
                     DNDEIQRNLNDTNYNRMWEYNNRGVGSKVVAEAKK"
     promoter        complement(1507. .1535)
                     /note="factor Sigma70; predicted +1 start at 806574"
     gene            complement(1582. .2058)
                     /gene="ybhB"
                     /note="b0773"
     CDS             complement(1582. .2058)
                     /gene="ybhB"
                     /function="orf; Unknown"
                     /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC73860.1"
                     /db_xref="GI:1786990"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDIERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     protein_bind    complement(2108. .2124)
                     /note="central position to predicted promoter:85.5"
                     /bound_moiety="RhaS predicted site"
     gene            complement(2117. .3406)
                     /gene="bioA"
                     /note="b0774"
     CDS             complement(2117. .3406)
                     /gene="bioA"
                     /EC_number="2.6.1.62"
                     /function="enzyme; Biosynthesis of cofactors, carriers:
                     Biotin"
                     /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
                     /codon_start=1
                     /transl_table=11
                     /product="7,8-diaminopelargonic acid synthetase"
                     /protein_id="AAC73861.1"
                     /db_xref="GI:1786991"
                     /translation="MTTDDLAFDQRHIWHPYTSMTSPLPVYPVVSAEGCELILSDGRR
                     LVDGMSSWWAAIHGYNHPQLNAAMKSQIDAMSHVMFGGITHAPAIELCRKLVAMTPQP
                     LECVFLADSGSVAVEVAMKMALQYWQAKGEARQRFLTFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRMDGEWDERDMVGFARLMAAHRHEIAAVIIEPIVQGAG
                     GMRMYHPEWLKRIRKICDREGILLIADEIATGFGRTGKLFACEHAEIAPDILCLGKAL
                     TGGTMTLSATLTTREVAETISNGEAGCFMHGPTFMGNPLACAAANASLAILESGDWQQ
```

```
                QVADIEVQLREQLAPARDAEMVADVRVLGAIGVVETTHPVNMAALQKFFVEQGVWIRP
                FGKLIYLMPPYIILPQQLQRLTAAVNRAVQDETFFCQ"
     promoter   complement(2193. .2221)
                /gene="bioA"
                /note="factor Sigma70; predicted +1 start at 807260"
     protein_bind 3411. .3450
                /note="central position to bioB promoter: -20"
                /bound_moiety="BioB documented site"
     protein_bind 3411. .3450
                /note="central position to predicted promoter: -20;genetic
                evidence for the site"
                /bound_moiety="BioB predicted site"
     protein_bind complement(3411. .3450)
                /note="central position to predicted promoter:50"
                /bound_moiety="BioB predicted site"
     protein_bind complement(3411. .3450)
                /note="central position to bioA promoter:50"
                /bound_moiety="BioB documented site"
     promoter   3413. .3441
                /note="factor Sigma70; promoter bioB; documented +1 at
                808525"
     promoter   complement(3447. .3473)
                /note="factor Sigma70; promoter bioA; documented +1 at
                808515"
     gene       3493. .4533
                /gene="bioB"
                /note="b0775"
     CDS        3493. .4533
                /gene="bioB"
                /EC_number="2.8.1.-"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
                /codon_start=1
                /transl_table=11
                /product="biotin synthesis, sulfur insertion?"
                /protein_id="AAC73862.1"
                /db_xref="GI:1786992"
                /translation="MAHRPRWTLSQVTELFEKPLLDLLFEAQQVHRQHFDPRQVQVST
                LLSIKTGACPEDCKYCPQSSRYKTGLEAERLMEVEQVLESARKAKAAGSTRFCMGAAW
                KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLSESQAQRLANAGLDYYNHNLDTSPEF
                YGNIITTRTYQERLDTLEKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSYVRLSAGREQMNEQTQAMC
                FMAGANSIFYGCKLLTTPNPEEDKDLQLFRKLGLNPQQTAVLAGDNEQQQRLEQALMT
                PDTDEYYNAAAL"
     gene       4530. .5684
                /gene="bioF"
                /note="b0776"
     CDS        4530. .5684
                /gene="bioF"
                /EC_number="2.3.1.47"
                /function="enzyme; Biosynthesis of cofactors, carriers:
                Biotin"
                /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
                /codon_start=1
                /transl_table=11
                /product="8-amino-7-oxononanoate synthase"
                /protein_id="AAC73863.1"
                /db_xref="GI:1786993"
                /translation="MSWQEKINAALDARRAADALRRRYPVAQGAGRWLVADDRQYLNF
                SSNDYLGLSHHPQIIRAWQQGAEQFGIGSGGSGHVSGYSVVHQALEEELAEWLGYSRA
                LLFISGFAANQAVIAAMMAKEDRIAADRLSHASLLEAASLSPSQLRRFAHNDVTHLAR
                LLASPCPGQQMVVTEGVFSMDGDSAPLAEIQQVTQQHNGWLMVDDAHGTGVIGEQGRG
                SCWLQKVKPELLVVTFGKGFGVSGAAVLCSSTVADYLLQFARHLIYSTSMPPAQAQAL

  Query Match          100.0%;  Score 25;  DB 1;  Length 11022;
  Best Local Similarity  100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db    3555 GGATCTGCTGTTTGAAGCGCAGCAG 3579
```

```
RESULT  13
AE004192
LOCUS       AE004192    12891 bp    DNA             BCT       31-JUL-2000
DEFINITION  Vibrio cholerae chromosome I, section 100 of 251 of the complete
            chromosome.
ACCESSION   AE004192 AE003852
VERSION     AE004192.1  GI:9655581
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 12891)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C. and White,O.
  TITLE     DNA sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
  JOURNAL   Nature 406 (6795), 477-483 (2000)
  MEDLINE   20406833
REFERENCE   2  (bases 1 to 12891)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12891
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /serotype="O1"
                     /db_xref="taxon:666"
                     /chromosome="I"
                     /note="biotype: El Tor"
     gene            complement(84. .1370)
                     /gene="VC1111"
     CDS             complement(84. .1370)
                     /gene="VC1111"
                     /note="similar to GB:J04423 SP:P12995 PID:457106
                     PID:490217 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="adenosylmethionine-8-amino-7-oxononanoate
                     aminotransferase"
                     /protein_id="AAF94270.1"
                     /db_xref="GI:9655582"
                     /translation="MMMDLAFDRCHIWHPYTSTLTPLTCYPVSRADGVMITLEDGRSL
                     VDGMSSWWAAIHGYNHPKLNQAAHQQIEQMSHIMFGGLTHQPAIELCQKLLKLAPNNL
                     QQVFLADSGSVAVEVSLKMALQYWHSKGQPRAKFLTLRHGYHGDTFAAMSVTDPDNSM
                     HSLYKGFLPEHIFANSPEGGFFDAWDERDIADFRHKLTEHHHQIAAVILEPIVQGAGG
                     MRIYHPEFLRQVRALCDEFGVLLILDEIATGFGRTGKLFACEHAGIQPDILCVGKALT
                     GGYMTLSATLTTQQVADTVCAGEAGCFMHGPTFMGNPLACAVASASLSLIEQGDWQQQ
                     VANIEAFFAEQLPKLNDSPRVKQTRWLGAIGVVETHLPVNMEVIQALFVEHGVWIRPF
                     GRLIYLMPPYISQPEHLQKLVDAMAAALNQPDCFKQ"
     gene            1497. .2549
                     /gene="VC1112"
     CDS             1497. .2549
                     /gene="VC1112"
                     /note="similar to GB:J04423 SP:P12996 PID:145425
                     PID:490219 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase"
```

```
                     /protein_id="AAF94271.1"
                     /db_xref="GI:9655583"
                     /translation="MEVRHNWTVAEVKALLDKPFMDLLFEAQQVHRLHHPHNHVQVST
                     LLSIKTGACPEDCKYCPQSAHYRTDVDKERLMEVERVLDAAQKAKNSGSTRFCMGAAW
                     KNPKERDMPLLKEMIRGVKDMGLETCMTLGMLTPDQAQQLAQAGLDYYNHNLDTSPEF
                     YGNIITTRTYQDRLDTLSHVRDAGMKICSGGIIGMGESTNDRAGLLVELANLPTHPES
                     VPINMLVKVKGTPLEQVDDVEPFDFVRLIAVARIMMPKSAVRLSAGREKMNEQMQALC
                     FMAGANSIFYGCKLLTTPNPAEDSDMLLFKKLGINREQVAQKPDEITENELLDRVVER
                     VAARPTASDLFYDAAL"
gene                 2533. .3687
                     /gene="VC1113"
CDS                  2533. .3687
                     /gene="VC1113"
                     /note="similar to GB:J04423 SP:P12998 PID:145426
                     PID:490225 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="8-amino-7-oxononanoate synthase"
                     /protein_id="AAF94272.1"
                     /db_xref="GI:9655584"
                     /translation="MTLRFKSRIEQALNEREHSRLTRRLQLLDRSAQGELIQQDRTYL
                     NFSSNDYLGLANDAELVQAWQTGLARFGAGSGASPMVTGFSSAHAELEHELCHWLGYE
                     RAVLFSSGFSANQALLFTLLEKEDLLLQDKLNHASLMEAGMLSPATMKRFKHNDTEHL
                     RQLLHEQSNSLVVTEGVFSMDGDCAPLAQIATLVKQRAWLMVDDAHGIGVLGEDGAGS
                     CQAAGIHPEILVVTFGKAFGLAGAAVLCDAHVGDYLTQFARHHVYSTAMPPAQAHALT
                     HAARMIQSQSWRREQLAELLACFDEQCRNIPGFVATQTPIKPWLLGSSDSALSASHEL
                     KQQGIWVSAIRPPTVPVGSARLRITITAAHTQAQIRRLSEQLKHVMEKDT"
gene                 3549. .4487
                     /gene="VC1114"
CDS                  3549. .4487
                     /gene="VC1114"
                     /note="similar to GB:J04423 SP:P12999 PID:145427
                     PID:490221 GB:U00096; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthesis protein BioC"
                     /protein_id="AAF94273.1"
                     /db_xref="GI:9655585"
                     /translation="MGECHSTTHRACRQCAFTYHDNGGAYPSADTPLIGAVKTCDGEG
                     HMSMTATELCVELKDKSAIAQAFGKAAAHYDQHAAFQRDVGLRLLQKMPSCLKGLRVL
                     DLGCGTGYFSALLRERGAQVVCADISHAMLEQAKQRCGDEGMSYQLADAEQLPFASAC
                     FDMVFSSLALQWCEDLSLPLSEIRRVLKPHGQAFLSTLLDGSLFELEQAWRSVDHHRH
                     INQFISINQVKIALAQAGCSQHHLDLAAITVWYEAAFMLMRDLKGIGANHVSGRSTGL
                     ISRRTLAKVELAYQSFRNQQGLVPATYQVCLGVIHR"
gene                 4463. .5179
                     /gene="VC1115"
CDS                  4463. .5179
                     /gene="VC1115"
                     /note="similar to GB:S68059 SP:P13000 PID:490223
                     PID:145428 PID:460901; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthetase"
                     /protein_id="AAF94274.1"
                     /db_xref="GI:9655586"
                     /translation="MFRGYSPMRNAIFIAGTDTDVGKTVASKAILQALATHNIATIGY
                     KPVAAGSDKTEFGYRNSDALHLMKAATVDMPYEDVNPYALVLPTSPHIAAKHENVTID
                     YALLSNKLSKHKQNAELVVVEGAGGWRVPTSDSDCLSTWVKQERLPVILVVGIKLGCL
                     SHAILTAEAIRADGLELVGWIANRINPGTEHYAEIIEHLEGRLGTPKLGEIPYMPKAK
                     RQELGKFIQLDHLLEPDTVA"
gene                 complement(5261. .5470)
                     /gene="VC1116"
CDS                  complement(5261. .5470)
                     /gene="VC1116"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF94275.1"
                     /db_xref="GI:9655587"
                     /translation="MSRRNLTTNKGKHMAEAHSYQMAIDLLCCHLGISEQEARTQLGL
                     EAKEQTLQQQIVDTQHSLMGLSAEK"
gene                 5697. .6560
                     /gene="VC1117"
CDS                  5697. .6560
                     /gene="VC1117"
                     /note="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
                     PID:1736470; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="heat shock protein HtpX"
                     /protein_id="AAF94276.1"
                     /db_xref="GI:9655588"
                     /translation="MKRILLFLATNLAVVLVLSVVLNIVYAVTGMQPGSLSGLLVMAA
                     VFGFGGAFISLLMSKSMALRSVGGVVIDTPRNEMEHWLLETVRRQANQAGIGMPTVAI
                     YDAPDMNAFATGAKRDDSLVAVSTGLLHNMTRDEAEAVLAHEVSHIANGDMVTMTLMQ
                     GVVNTFVIFLSRFIANIVASRDSEEGEGSNMMVYFGVSMVLELVFGFLASFITMWYSR
                     HREFHADAGAAQLVGKHKMIAALERLKMGQESHLEGSMMAFGITGKRSLSELMMTHPP
                     LEKRIAALRNM"
gene                 6770. .7198
                     /gene="VC1118"
CDS                  6770. .7198
                     /gene="VC1118"
                     /note="similar to PID:1653608; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, putative"
                     /protein_id="AAF94277.1"
                     /db_xref="GI:9655589"
                     /translation="MSIMDVQHVAQFYQQLDKSQLHRLIEIYHPDVVFEDAAHRIEGF
                     DALYQYFLNLYQNVHTCTFTIHEQYAVNEGAFLVWTMHLRHPKLAKGEQVDVKGVSHL
                     HFAEGKVTYHRDYFDMGEMLYEQLPVLGQVIRAIKRRLGQ"
gene                 7195. .7920
                     /gene="VC1119"
CDS                  7195. .7920
                     /gene="VC1119"
                     /note="similar to SP:P25970; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="oxidoreductase, short-chain
                     dehydrogenase/reductase family"
                     /protein_id="AAF94278.1"
                     /db_xref="GI:9655590"
                     /translation="MTAVFITGATSGIGKQLALDYAKQGWQVIACGRNQSVLDSLHTQ
                     YANIFPLAFDVTDHPGTKAALAQLPCQPELWILNAGDCEYIDDGKMDVTLMARVFNIN
                     VLGVAAVIEGIQPHLSCGHRVAIVGSIASELALPRAEAYGASKAAVAYLARTLQLDWR
                     PLGIEVTTIFPGFVATPLTDRNTFAMPMIITVERAAQEIKAGLARGVSQLYFPKRFTW
                     LIRLLGALPYAWQGRLVRRLLKA"
gene                 7931. .9211
                     /gene="VC1120"
CDS                  7931. .9211
                     /gene="VC1120"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11

  Query Match          100.0%;  Score 25;  DB 1;  Length 12891;
  Best Local Similarity  100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db    1559 GGATCTGCTGTTTGAAGCGCAGCAG 1583


RESULT  14
AE005258
LOCUS       AE005258    13501 bp    DNA             BCT       21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
            of 155.
ACCESSION   AE005258 AE005174
```

```
VERSION     AE005258.1  GI:12513751
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 13501)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .13501
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .7576
                     /note="O-island #36; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655; Cryptic prophage
                     CP-933K; includes one copy of the 13 bp direct repeat that
                     flanks the prophage"
     gene            66. .665
                     /gene="lomK"
                     /note="Z0981"
     CDS             66. .665
                     /gene="lomK"
                     /function="putative membrane; Other or unknown (Phage or
                     Prophage Related)"
                     /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                     residues 1 to 199 of 199 from GenPept 118 :
                     gi|7532789|gb|AAF63231.1|AF151091_2 (AF151091) Lom
                     [prophage P-EibA]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative outer membrane protein of prophage
                     CP-933K"
                     /protein_id="AAG55137.1"
                     /db_xref="GI:12513752"
                     /translation="MRKVCAAILSAAICLAVSGVPAWASEHQSTLSAGYLHASTDAPG
                     SDDLNGINVKYRYEFTDTLGLITSFSYANAEDEQKTHYSDTRWHEDYVRNRWFSVMAG
                     PSVRVNEWFSAYAMAGVAYSRVSTFSGDYFRVTDNKRKTHDVLTGSDDARYSNTSLAW
                     GAGVQFNPTESVAVDVAYEXSGSGDWRTDGFIVGVGYKF"
     gene            719. .2041
                     /gene="Z0982"
     CDS             719. .2041
                     /gene="Z0982"
                     /function="putative structure; Structural component (Phage
                     or Prophage Related)"
                     /note="Residues 164 to 440 of 440 are 68.79 pct identical
                     to residues 381 to 645 of 645 from GenPept 118 :
                     gi|4585436|gb|AAD25464.1|AF125520_59 (AF125520) putative
                     tail fiber protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative tail component of prophage CP-933K"
                     /protein_id="AAG55138.1"
                     /db_xref="GI:12513753"
                     /translation="MGMAAVQISGVLKDGAGKPIQNCTIQLKARRNSTXVVVNTVASE
                     NPDEAGRYSMDVEYGQYSVTLLVEGFPPSHAGTISVYEDSQPGTLNDFLGAMTEDDAR
                     PEALRRFEQMVEEAARHAEEAKKNAGEAETSARNAGISASKAEASAANADTSAEDASE
                     SARQAAESAASAKKSEEASSSSASEAAQKASESLQSATDAELSKKTAESAAGNAARDA
                     TTSTEKARESAESAQSAEQSRIAAEDAVNRIPTVVGPPGPKGEPGPAGPQGPKGDKGE
                     RGDTGPAGATGERGPGGDTGPAGPQGPKGDRGERGETGLTGNAGPQGPKGDTGAAGPA
                     GPQGPKGETGAAGPVGATGPQGPKGDPGETQIRFRLGPMRIIETNSYGWFPGTDGALI
                     TGLTFLDPKDATQVQGMFQHLQVRFGDGPWQDVKGLDEVGSDTGRTGE"
     gene            2004. .2312
                     /gene="Z0984"
     CDS             2004. .2312
                     /gene="Z0984"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 1 to 102 of 102 are 98.03 pct identical to
                     residues 28 to 129 of 129 from GenPept 118 :
                     gi|4585437|gb|AAD25465.1|AF125520_60 (AF125520)
                     hypothetical protein [Bacteriophage 933W]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55139.1"
                     /db_xref="GI:12513754"
                     /translation="MKWAVIQAELENDMNILRKLMQSLCGCGKHDDCENGRSLTAQLR
                     LGPADILESDENGIIPEQDRVITQVVILDADKKQIQCVVRPLQILRADGTWENIGGMK
                     "
     gene            2489. .3469
                     /gene="Z0985"
     CDS             2489. .3469
                     /gene="Z0985"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 13 to 326 of 326 are 55.73 pct identical
                     to residues 18 to 331 of 336 from GenPept 118 :
                     gi|6960367|gb|AAF33527.1| (AF170176) hypothetical protein
                     predicted by Glimmer [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55140.1"
                     /db_xref="GI:12513755"
                     /translation="MLSPIRTTFHNSVNIVQSSPCQTVSFAGKEYELKVIDEKTPILF
                     QWFEPNPERYKKDEVPIVNTKQHPYLDNVTNAARIESDRMIGIFVDGDFSVNQKTAFS
                     KLERDFENVMIIYREDVDFSMYDRKLSDIYHDIICEQRLRTEDKRDEYLLNLLEKELR
                     EISKAQDSLISMYAKKRNHAWFDFFRNLALLKAGEIFRCTYNTKNHGISFGEGCIYLD
                     MDMILTGKLGTIYAPDGISMHVDRRNDSVNIENSAIIVNRSNHPALLEGLSFMHSKVD
                     AHPYYDGLGKGVKKYFNFTPLHNYNHFCDFIEFNHPNIIMNTSQYTCSSW"
     gene            3530. .4522
                     /gene="Z0986"
     CDS             3530. .4522
                     /gene="Z0986"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55141.1"
                     /db_xref="GI:12513756"
                     /translation="MKIPSLQSNFNFSAPAGYSAPIAPNRAENAYADYVLDIGKRIPL
                     SAADLSNVYESVIRAVHDSRSRLIDQHTVDMIGNTVLDALSRSQTFRDAVSYGIHNEK
                     VHIGCIKYRNEYELNEESSVKIDDIQSLTCNELYEYDVGQEPIFPICEAGENDNEEPY
                     VSFSVAPDTDSYEMPSWQEGLIHEIIHHVTGSSDPSGDSNIELGPTEILARRVAQELG
                     WSVPDFKGYAEPEREAHLRLRNLNALRQAAMRHEENERAFFERLGTISDRYEASPDFT
                     EYSAVSNIGYGFIQQHDFPGLAINDNLQDANQIQLYHGAPYIFTFGDVDKHNQQ"
     gene            5350. .6231
                     /gene="Z0989"
     CDS             5350. .6231
                     /gene="Z0989"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
```

```
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55142.1"
                     /db_xref="GI:12513757"
                     /translation="MLSPYSVNLGCSWNSLTRNLTSPDNRVLSSVRDAAVHSDNGAQV
                     KVGNRTYRVVATDNKFCVTRESHSGCFTNLLHRLGWPKGEISRKIEVMLNASPVSAAM
                     ERGIVHSNRPDLPPVDYAPPELPSVDYNRLSVPGNVIGKGGNAVVYEDAEDATKVLKM
                     FTTSQSNEEVTSEVRCFNQYYGAGSAEKIYGNNGDIIGIRMDKINGESLLNISSLPAQ
                     AEHAIYDMFDRLEQKGILFVDTTETNVLYDRAKNEFNPIDISSYNVSDRSWSESQIMQ
                     SYHGGKQDLISVVLSKI"
     gene            6462. .7160
                     /gene="Z0990"
     CDS             6462. .7160
                     /gene="Z0990"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933K"
                     /protein_id="AAG55143.1"
                     /db_xref="GI:12513758"
                     /translation="MRPTSLNLVLHQSSRSSSMSDTDIESLVKASSVQWIKNNPQLRF
                     QGTDHNIYQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRD
                     IDAENHRGTGSDFHCNLNAVEYPCGEGISVVDFHATIVFHELLHVFHNLNGERLKVES
                     SRAESQKYSPLLLEEARTVGLGAFSEEVLSENKFHEEIGMPRRTSYPXDSALIHDDNT
                     VSLGFQQVRLHPLL"
     gene            complement(7668. .8144)
                     /gene="ybhB"
                     /note="Z0992"
     CDS             complement(7668. .8144)
                     /gene="ybhB"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 158 of 158 are 99.36 pct identical to
                     residues 1 to 158 of 158 from Escherichia coli K-12 Strain
                     MG1655: B0773"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG55144.1"
                     /db_xref="GI:12513759"
                     /translation="MKLISNDLRDGDKLPHRHVFNGMGYDGDNISPHLAWDDVPAGTK
                     SFVVTCYDPDAPTGSGWWHWVVVNLPADTRVLPQGFGSGLVAMPDGVLQTRTDFGKTG
                     YDGAAPPKGETHRYIFTVHALDVERIDVDEGASGAMVGFNVHFHSLASASITAMFS"
     gene            complement(8203. .9492)
                     /gene="bioA"
                     /note="Z0993"
     CDS             complement(8203. .9492)

  Query Match             100.0%;  Score 25;  DB 1;  Length 13501;
  Best Local Similarity   100.0%;  Pred. No. 0.00013;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db    9641 GGATCTGCTGTTTGAAGCGCAGCAG 9665


RESULT  15
AP002553
LOCUS       AP002553   297816 bp    DNA             BCT       07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION   AP002553 BA000007
VERSION     AP002553.1  GI:13360211
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 297816)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .297816
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            79. .1245
                     /gene="ECs0753"
     CDS             79. .1245
                     /gene="ECs0753"
                     /note="similar to SUCC_ECOLI gi|1786948 percent identity
                     100 in 388 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="succinyl-CoA synthetase beta subunit"
                     /protein_id="BAB34176.1"
                     /db_xref="GI:13360212"
                     /translation="MNLHEYQAKQLFARYGLPAPVGYACTTPREAEEAASKIGAGPWV
                     VKCQVHAGGRGKAGGVKVVNSKEDIRAFAENWLGKRLVTYQTDANGQPVNQILVEAAT
                     DIAKELYLGAVVDRSSRRVVFMASTEGGVEIEKVAEETPHLIHKVALDPLTGPMPYQG
                     RELAFKLGLEGKLVQQFTKIFMGLATIFLERDLALIEINPLVITKQGDLICLDGKLGA
                     DGNALFRQPDLREMRDQSQEDPREAQAAQWELNYVALDGNIGCMVNGAGLAMGTMDIV
                     KLHGGEPANFLDVGGGATKERVTEAFKIILSDDKVKAVLVNIFGGIVRCDLIADGIIG
                     AVAEVGVNVPVVVRLEGNNAELGAKKLADSGLNIIAAKGLTDAAQQVVAAVEGK"
```

```
gene            1245. .2114
                /gene="ECs0754"
CDS             1245. .2114
                /gene="ECs0754"
                /note="similar to SUCD_ECOLI gi|1786949 percent identity
                100 in 289 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="succinyl-CoA synthetase alpha subunit"
                /protein_id="BAB34177.1"
                /db_xref="GI:13360213"
                /translation="MSILIDKNTKVICQGFTGSQGTFHSEQAIAYGTKMVGGVTPGKG
                GTTHLGLPVFNTVREAVAATGATASVIYVPAPFCKDSILEAIDAGIKLIITITEGIPT
                LDMLTVKVKLDEAGVRMIGPNCPGVITPGECKIGIQPGHIHKPGKVGIVSRSGTLTYE
                AVKQTTDYGFGQSTCVGIGGDPIPGSNFIDILEMFEKDPQTEAIVMIGEIGGSAEEEA
                AAYIKEHVTKPVVGYIAGVTAPKGKRMGHAGAIIAGGKGTADEKFAALEAAGVKTVRS
                LADIGEALKTVLK"
gene            2781. .3686
                /gene="ECs0755"
CDS             2781. .3686
                /gene="ECs0755"
                /note="probable transcriptional regulator,similar to
                transcriptional regulators e.g. glycine cleavage system
                transcription activator (gcv operon activator) -
                Escherichia coli gi|417043|sp|P32064|GCVA_ECOLI percent
                identity 31 in 300 aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transcription regulator"
                /protein_id="BAB34178.1"
                /db_xref="GI:13360214"
                /translation="MRGKIPKTELLVTFEVVARHESYTRAAEELALTQSAVFRQVSAL
                EEFLRTPLFSHSKKRIFLNDTGKYYLGIVKETLNKLERDTNTIMTWQPTVQVIELAVN
                PTFSTHWLIPNLHEFTKLHPDIIVNIHSLANNGDFLNREYDAVIMRENFCAPWAEVEY
                LFEEEILPVCSGSLLAMSDQKLSVAELLTELPLLHQSTRITGWEEWFALSGVSSPLVN
                NGPRFDLLSMLIAAVRSNLGVALLPRFAIQHDLDSGDMVIPCDVPIRTGNRFIMTWQE
                EKSDSPHLQQFREWLLAKSVVPQEM"
gene            complement(3720. .4322)
                /gene="ECs0756"
CDS             complement(3720. .4322)
                /gene="ECs0756"
                /note="probable cob(I)alamin adenosyltransferase,similar
                to cob(I)alamin adenosyltransferases (corrinoid
                adenosyltransferases) e.g. [Escherichia coli]
                gi|115148|sp|P13040|BTUR_ECOLI percent identity 67 in 200
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative cob(I)alamin adenosyltransferase"
                /protein_id="BAB34179.1"
                /db_xref="GI:13360215"
                /translation="MEARISTERHVQRQQKLKEQVDTRVAAAIEKKGILIVFTGNGKG
                KSTAAFGTVTRAVGHGKTVGVAQYIKGQWDNGEYNLLQPLGVEFHIMGTGFTWETQNR
                QADIDAAKEVWSESKRMLADKRYDLVVLDELTYMLAYHYLDTEEVIASLQNRPAQQSV
                IVTGRGCHSQILKMADTVSEIRPVKHAFDNGIQAQPGIDW"
gene            complement(4332. .5984)
                /gene="ECs0757"
CDS             complement(4332. .5984)
                /gene="ECs0757"
                /note="probable fumarate hydratase,similar to fumarate
                hydratases e.g. fumarate hydratase class I, aerobic
                (fumarase) - Escherichia coli
                gi|120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
                aa"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative fumarate hydratase"
                /protein_id="BAB34180.1"
                /db_xref="GI:13360216"
                /translation="MSKPFIWQEPFLQNKDGTEYTLISDQHITVTELDGEEVIKIAPE
                ALTLLARQAFYEASFFLRSAHLQQVASILNDQQASSNDKYVALQLLRNAEVSAKGVLP
                NCQDTGTATIVASKGQQIWTGGNDAEALSKGIYSTFQENNLRFSQNAPLDMYTEVNTQ
                TNLPAQIDISAVAGDEYHFLCVNKGGGSANKAALYQETKSLLQPEKLTAFLIEKMKSL
                GTAACPPYHIAFVVGGLSADQTLKIAKLASTKYYDNLPTSGNEQGQAFRDIELEKVLL
                EASQQFGIGAQFGGKYFAHDIRVIRLPRHGGSCPIAMALSCSADRNIKAKINKHGIWL
                EKLEHNPGQYIPASLREENHAQHVQLDLNRPLRDVMQDLARLPVGTRVSLSGPIVVAR
                DIAHAKIKARLDSGEPMPEYLKHHIVYYAGPAKTPENMACGSLGPTTGGRMDGYIDTF
                QAAGGSLVMLSKGNRSQQVTDACHKHGGFNLGSIGGAAALLAQEYVKSLRCLEYPELG
                MEAVWMMEVENLPAFILVDDKGNNFFSQFEQQHRCASCPAGH"
gene            complement(6092. .7372)
                /gene="ECs0758"
CDS             complement(6092. .7372)
                /gene="ECs0758"
                /note="probable transport protein,similar to
                glutamate/aspartate transport proteins (proton glutamate
                symport proteins) e.g. [Bacillus stearothermophilus]
                gi|121467|sp|P24943|GLTT_BACST percent identity 38 in 416
                aa,also similar to C4-dicarboxylate transpor"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="putative transport protein"
                /protein_id="BAB34181.1"
                /db_xref="GI:13360217"
                /translation="MKKISLTTMILLALVLGMIIGVVLNNTASPETAKLYAQEISIFT
                TIFLRLIKMIIAPLVVSTLVVGIAKMGDAKALGRIFSKTLFLFICASLLSIALGLITV
                NFFMPGTGINFVAHGAETTGVVAAEPFTLKVFISHAFPTSIVDAMAHNEILQIVVFSI
                FLGCSLTAIGEKGSAIVHALDSLAHAMLKLTGYVMLFAPLTVFAAISALIAERGLAVM
                VSAGIFMGEFYFTMLLLWVLLIGLAIVYVGPCIRRLTRALSEPALLAFTTSSSEAAFP
                GTLEKLEQFGVSPKIASFVLPIGYSFNLVGSMAYCSFATVFIAQACNIHLSIGEQITM
                LLILMLTSKGMAGVPRASMVVIAATLNQFNIPEAGLILLMGVDPFLDMGRSATNVMSN
                AMGAAMVSRWEGEHFGEGCRGKALKPNESNVALP"
gene            complement(7533. .7850)
                /gene="ECs0759"
CDS             complement(7533. .7850)
                /gene="ECs0759"
                /note="unknown"
                /codon_start=1
                /transl_table=11
                /evidence=not_experimental
                /product="hypothetical protein"
                /protein_id="BAB34182.1"
```

```
  Query Match             100.0%;  Score 25;  DB 1;  Length 297816;
  Best Local Similarity   100.0%;  Pred. No. 0.00014;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagcgcagcag 25
          |||||||||||||||||||||||||
Db  92373 GGATCTGCTGTTTGAAGCGCAGCAG 92397


RESULT  16
ECOBIOB
LOCUS       ECOBIOB       128 bp    DNA             BCT       03-JAN-1995
DEFINITION  Escherichia coli biotin (bioB) gene, early terminator region.
ACCESSION   M27731
VERSION     M27731.1  GI:341755
KEYWORDS    bioB gene; biotin.
SOURCE      Escherichia coli (strain K-12) DNA.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 128)
  AUTHORS   Nath,S.K.
  TITLE     Attenuation of transcription of biotin genes in Escherichia coli
  JOURNAL   Can. J. Microbiol. 34 (12), 1288-1296 (1988)
  MEDLINE   89167942
FEATURES             Location/Qualifiers
     source          1. .128
                     /organism="Escherichia coli"
                     /strain="K-12"
```

```
                     /db_xref="taxon:562"
     gene            4. .128
                     /gene="bioB"
     -10_signal      4. .9
                     /gene="bioB"
     mRNA            20. .>128
                     /gene="bioB"
     terminator      103. .121
                     /gene="bioB"
                     /note="early terminator"
BASE COUNT       31 a     28 c     34 g     35 t
ORIGIN


  Query Match             84.0%;  Score 21;  DB 1;  Length 128;
  Best Local Similarity   100.0%;  Pred. No. 0.03;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgca 21
           |||||||||||||||||||||
Db      24 GGATCTGCTGTTTGAAGCGCA 44


RESULT  17
AF248314
LOCUS       AF248314     8227 bp    DNA             BCT       24-JAN-2001
DEFINITION  Uncultured bacterium pCosAS1 urocanase-like protein (hutU) gene,
            partial cds; histidine ammonia-lyase-like protein (hutH), DAPA
            aminotransferase BioA (bioA), biotin synthase BioB (bioB), 7-KAPA
            synthetase (bioF), biotin biosynthesis BioC-like protein (bioC),
            and dethiobiotin synthase BioD (bioD) genes, complete cds; ABC
            transporter-like protein (elsA) gene, partial cds; and unknown
            gene.
ACCESSION   AF248314
VERSION     AF248314.1  GI:12407610
KEYWORDS    .
SOURCE      uncultured bacterium pCosAS1.
  ORGANISM  uncultured bacterium pCosAS1
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 8227)
  AUTHORS   Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streit,W.R.
  TITLE     Direct Cloning from Enrichment Cultures, a Reliable Strategy for
            Isolation of Complete Operons and Genes from Microbial Consortia
  JOURNAL   Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
   PUBMED   11133432
REFERENCE   2  (bases 1 to 8227)
  AUTHORS   Entcheva,P., Liebl,W. and Streit,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Mikrobiologie und Genetik, Universitaet
            Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES             Location/Qualifiers
     source          1. .8227
                     /organism="uncultured bacterium pCosAS1"
                     /db_xref="taxon:143793"
                     /clone="pCosAS1"
                     /note="unknown organism, cosmid clone derived from
                     environmental consortium"
     gene            <1. .502
                     /gene="hutU"
     CDS             <1. .502
                     /gene="hutU"
                     /note="putative urocanate hydratase"
                     /codon_start=2
                     /transl_table=11
                     /product="urocanase-like protein"
                     /protein_id="AAG53593.1"
                     /db_xref="GI:12407618"
                     /translation="PRERISFQGLPARICWVGLEWRQKLGLAFNEMVRRGEVSAPIVI
                     GRDHLDSGSVASPNRETEAMQDGSDAVSDWPLLNALLNTASGATWVSLHHGGGVGMGF
                     SQHAGMVIVCDGTDDAAARIARVLHNDPATGVMRHADAGYDIAIDCAKEQGLNLPMIA
                     ATQGQH"
     gene            502. .2031
                     /gene="hutH"
     CDS             502. .2031
                     /gene="hutH"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="histidine ammonia-lyase-like protein"
                     /protein_id="AAG53586.1"
                     /db_xref="GI:12407611"
                     /translation="MNALTLTPGTLTLAQLRQVWQQPLQLTLDESAHEAINDSVACVE
                     AIVAEGRTAYGINTGFGLLAQTRIATHDLENLQRSLVLSHAAGVGEPLDDDIVRLMMV
                     LKINSLARGFSGIRLSVIQALIALVNAGVYSVDPAKGSVGASGDLAPLAHMSLTLLGE
                     GKARYRGEWLPAATALQKAGLAPVTLAAKEGLALLNGTQASTAFALRGLFEAEDLFAS
                     AVVCGALTTEAVLGSRRPFDARIHEVRGQRGQIDAAALFRHVLTDTSAIAESHHNCDK
                     VQDPYSLRCQPQVMGACLTQMRQVAEVLLVESNAVSDNPLVFAAENEMVFRGNFHAEP
                     VAMAADNLALAIAEIGALSERRIALMMDKHMSQLPPFLVRNGGVNSGFMIAQVTAAAL
                     ASENKGLCHPHTTSVDKPCRPSANQEDHVSMAPAAGRRLWEMAGNTRGVLAVEWLAAC
                     QGADLRDGLTSSPLLEQARQSCGEQVAHYDDDRFFAPDIEAAISLLNKGSLVGLLPAF
                     L"
     CDS             complement(2130. .2606)
                     /note="ORF1; similar to hypothetical 17.1 kDa protein in
                     modC-bioA intergenic region"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAG53587.1"
                     /db_xref="GI:12407612"
                     /translation="MKLISQDLRDGDKLPLRHVFNGMGYEGDNISPHLAWDEVPSGTK
                     SFVVTCYDPDAPTDSGWWHWIVANIPADTRVLPQGSGSSLVALPEGAVETRTDFGKAG
                     YGGAAPPKGETHRYIFTVHALNVERIEVDEEASGAMVGFNVHFHTLGSASITAMYS"
     gene            complement(2651. .3940)
                     /gene="bioA"
     CDS             complement(2651. .3940)
                     /gene="bioA"
                     /note="7,8-diamino-pelargonic acid aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="DAPA aminotransferase BioA"
                     /protein_id="AAG53588.1"
                     /db_xref="GI:12407613"
                     /translation="MTTDDLAFDRQHIWHPYTSMTQPLPVYPVVEAHGCELTLASGEQ
                     LVDGMSSWWAAIHGYNHPRLNAAMKAQIDRMSHVMFGGITHQPAVDLCRRLVAMTPET
                     LECVFPADSGSVAVEVAMKMALQYWHAKGESRQRFITFRNGYHGDTFGAMSVCDPDNS
                     MHSLWKGYLPENLFAPAPQSRFDGAWDEMDMVGLAPLMAAHRHEIAAVILEPIVQGAG
                     GMRMYHPEWLKRIRRMCDREGILLIADEIATGFGRTGKLFACEHAGITPDILCLGKAL
                     TGGTMTMSATLTTRQVAETISNGEAGCFMHGPTFMGNPLACAVASESLALLESGEWQD
                     QVAAIESQLQEELAVARDSEFVADVRVLGAIGVVETTHPVNMAALQRFFVARGVWVRP
                     FGKLIYLMPPYVIRPDQLSKLTRAVVDAVQAPAHFTR"
     gene            4029. .5069
                     /gene="bioB"
     CDS             4029. .5069
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthase BioB"
                     /protein_id="AAG53589.1"
                     /db_xref="GI:12407614"
                     /translation="MAHQARWTLSQVTALFEKPLLELLFEAQQIHRQHFDPQQIQVST
                     LLSIKTGACPEDCKYCPQSARYKTGLESERLMEVEQVLESARQAKNAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLQACMTLGTLDDHQAQRLASAGLDYYNHNLDTFPEF
                     YGNIITTRTYQERLDTLDKVRDAGIKVCSGGIVGLCETVTDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVXAFDXIRTXSVAXVMXPTXLVRLSTGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDVQLFRKLGLNPQQTAVLTGDNEQQHQLEQQLFN
                     ADTDQFYNAATV"
     gene            5066. .6223
                     /gene="bioF"
     CDS             5066. .6223
                     /gene="bioF"
                     /note="7-keto-8-amino-pelargonic acid synthetase; similar
                     to BioF"
                     /codon_start=1
                     /transl_table=11
                     /product="7-KAPA synthetase"
                     /protein_id="AAG53590.1"
```

```
                     /db_xref="GI:12407615"
                     /translation="MTWQQRIDAALDARRAADALRVRTPVENGAGRFLIREQRRYCNF
                     SGNDYLGLSQHPEIIRAWQQGAEQYGVGSGGSGHVSGYTHAHQALEEHLAGWLGYPRA
                     LLFISGFAANQALIAALTDKEDRIVADRLSHASLLEAASHSPAQLRRFAHNDVAQLDS
                     LLDKPCHGQQLVVTEGIFSMDGDSAPLGVIHAAAQRQGSWLLVDDAHGIGVTGPEGRG
                     SAASQGVKPELLVVTFGKGVGVSGAAVLCSDAVADYLLQFARHLIYSTSMPPAQAVAL
                     LAAFNVIRSDEDNARRQRLAALIDQFRAGVRDLPVQITASQSAIQPLIVGENARALHL
                     AQRLRERGIWVTAIRPPTVPAGTARLRLTLTAAHQPADITALVEALHVAGE"
     gene            6207. .6962
                     /gene="bioC"
     CDS             6207. .6962
                     /gene="bioC"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin biosynthesis BioC-like protein"
                     /protein_id="AAG53591.1"
                     /db_xref="GI:12407616"
                     /translation="MSQVNKQAIAAAFGRAAQSYSRHDALQRQSADGLRALLAENTFA
                     SVLDAGCGPGSNSLAWRQTGSVVTAMDLCAPMLEEARRNQAADSYLQADIEAIPLADA
                     QFSLAWSHLAVQXCASLPRALAELWRVTQLGGKVAFTTLLNGSLPELNQAWQAVDTQP
                     HANRFLSGQQVNDALSGWQYRCAVQTVTLLFDDALSAMHSLKGIGATHLHAGRAAKPL
                     TRGQLQQLAQAWPQQAGKFPLSYQLFHGIIERD"
     gene            6940. .7680
                     /gene="bioD"
     CDS             6940. .7680
                     /gene="bioD"
                     /codon_start=1
                     /transl_table=11
                     /product="dethiobiotin synthase BioD"
                     /protein_id="AAG53592.1"
                     /db_xref="GI:12407617"
                     /translation="MELSNVTERYFVTGTDTEVGKTVASTALLQAARALGRHTAGYKP
                     VASGSEMTAQGLRNSDALALQRNSSLALSYPAVNPYTFAEPTSPHIVSADEQRPIEFA
                     TLSAGLRALETQADWVLVEGAGGWFTPLSQEQTFADWVMAERLPVILVVGVKLGCINH
                     AMLTAQAIRQAGLHFAGWVANGVVVPGKRHAEYMATLRRVLSAPLLGEIPWLGEQADT
                     SAVGQYLDLTALTPSAPSSGQSAGHPAR"
     gene            complement(7619. .>8227)
                     /gene="elsA"
     CDS             complement(7619. .>8227)
                     /gene="elsA"
                     /note="putative urea/short-chain amide ABC transporter"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter-like protein"
                     /protein_id="AAG53594.1"
                     /db_xref="GI:12407619"
                     /translation="DCXLRTRXGENXLVSCIAGLLPIASGTMVWHEAPGAVNQAGTEI
                     XVVPQDRRIFSQLTVEENLHIAQMAKGIAGGSLSRDVFEFFPELYPLRQLRASGLNDD
                     QQYQLALANALVTRPRLLILDEPSRGTGQLFTVKLAQLLLRFNRELGLSVLLAEQHLS
                     LIRRVADRFCLLYRGRNVAQGRVSELDDPLIAHWMGPTASAQ"
BASE COUNT     1512 a    2642 c    2605 g    1453 t      15 others
ORIGIN


  Query Match             84.0%;  Score 21;  DB 1;  Length 8227;
  Best Local Similarity   100.0%;  Pred. No. 0.034;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||
Db    4095 CTGCTGTTTGAAGCGCAGCAG 4115


RESULT  18
EHBIOTOP2
LOCUS       EHBIOTOP2     1041 bp    DNA             BCT       04-NOV-1996
DEFINITION  Erwinia herbicola biotin synthetase (bioB) gene, complete cds.
ACCESSION   U38648
VERSION     U38648.1  GI:1228112
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      Erwinia herbicola strain=Eho10.
  ORGANISM  Pantoea agglomerans
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Pantoea.
REFERENCE   1  (bases 1 to 1041)
  AUTHORS   Wu,C.H., Chen,H.Y. and Shiuan,D.
  TITLE     Isolation and characterization of the Erwinia herbicola bio operon
            and the sequences of the bioA and bioB genes
  JOURNAL   Gene 174 (2), 251-258 (1996)
  MEDLINE   97045821
REFERENCE   2  (bases 1 to 1041)
  AUTHORS   Shiuan,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-1995) David Shiuan. Biology, Natl. Sun Yat-Sen
            University, Kaohsiung, Taiwan, ROC
FEATURES             Location/Qualifiers
     source          1. .1041
                     /organism="Pantoea agglomerans"
                     /strain="Eho10"
                     /db_xref="taxon:549"
     gene            1. .1041
                     /gene="bioB"
     CDS             1. .1041
                     /gene="bioB"
                     /codon_start=1
                     /transl_table=11
                     /product="biotin synthetase"
                     /protein_id="AAC44534.1"
                     /db_xref="GI:1228114"
                     /translation="MAHHARWTLSQVTELFDKPLLDLLFEAQTIHRQHFDPRQVQVST
                     LLSIKTGACPEDCKYCPQSSRYKTGLETERLMEVEQVLDSARKAKNAGSTRFCMGAAW
                     KNPHERDMPYLEQMVQGVKAMGLEACMTLGTLDETQAQRLASAGLDYYNHNLDTSPEF
                     YGNIITTRTYQERLDTLDKVRDAGIKVCSGGIVGLGETVKDRAGLLLQLANLPTPPES
                     VPINMLVKVKGTPLADNDDVDAFDFIRTIAVARIMMPTSFVRLSAGREQMNEQTQAMC
                     FMAGANSIFYGCKLLTTPNPEEDKDVQLFRKLGLNPQQTDVMTGDNEQQQKLEEQIFN
                     ADTDQFYNAAAL"
BASE COUNT      230 a    322 c    308 g    181 t
ORIGIN


  Query Match             72.0%;  Score 18;  DB 1;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 2.1;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ctgctgtttgaagcgcag 22
           ||||||||||||||||||
Db      67 CTGCTGTTTGAAGCGCAG 84


RESULT  19
E00893
LOCUS       E00893       1121 bp    DNA             PAT       29-SEP-1997
DEFINITION  Genomic DNA encoding biotin Synthetase.
ACCESSION   E00893
VERSION     E00893.1  GI:2169154
KEYWORDS    JP 1986149091-A/1.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Hirono,Y., Kojima,T. and Kimura,H. .
  TITLE     DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
            PRODUCTION OF BIOTIN
  JOURNAL   Patent: JP 1986149091-A 1 07-JUL-1986;
            NIPPON SODA CO LTD
COMMENT     OS   Escherichia coli
            PN   JP 1986149091-A/1
            PD   07-JUL-1986
            PF   24-DEC-1984 JP 1984272605
            PI   HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
            C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
            C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
```

```
            CC   hypothetical: No;
            CC   anti-sense: No;
            CC   *source: strain=Escherichia coli Ns101;
            CC   Feature is identified by experimental;
            FH   Key             Location/Qualifiers
            FH
            FT   CDS             42. .1079
            FT                   /product='biotin synthetase'.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      289 a    296 c    325 g    211 t
ORIGIN


  Query Match             72.0%;  Score 18;  DB 6;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 2.1;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagc 18
          ||||||||||||||||||
Db    104 GGATCTGCTGTTTGAAGC 121


RESULT  20
AF250347/c
LOCUS       AF250347     1044 bp    mRNA            VRT       01-JUL-2001
DEFINITION  Xenopus laevis transcription factor NKX-2.1 mRNA, complete cds.
ACCESSION   AF250347
VERSION     AF250347.1  GI:14578835
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1044)
  AUTHORS   Hollemann,T. and Pieler,T.
  TITLE     Xnkx-2.1: a homeobox gene expressed during early forebrain, lung
            and thyroid development in Xenopus laevis
  JOURNAL   Dev. Genes Evol. 210 (11), 579-581 (2000)
  MEDLINE   21110556
   PUBMED   11180810
REFERENCE   2  (bases 1 to 1044)
  AUTHORS   Hollemann,T. and Pieler,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2000) Entwicklungsbiochemie, University of
            Goettingen, Humboldtallee 23, Goettingen, LowerSaxony 37073,
            Germany
FEATURES             Location/Qualifiers
     source          1. .1044
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     primer_bind     1. .29
     CDS             1. .1044
                     /note="Xnkx-2.1"
                     /codon_start=1
                     /product="transcription factor NKX-2.1"
                     /protein_id="AAK69025.1"
                     /db_xref="GI:14578836"
                     /translation="MSMSPKHTTPFSVSDILSPLEESYKKVAMEGAGLGAPLTAAAYR
                     QSQVSQASMQQHGMGHNGPVSAAYHMTAAGVPQLSHTTMGGYCNGNLGNLGNMSELPP
                     YQDTMRNSSATGWYGANPDPRFSTISRFMGPSGGMNMEGLGNMGSLGDVGKSMTPLQA
                     TPRRKRRVLFSQAQVYELERRFKQQKYLSAPEREHLASMIHLTPTQVKIWFQNHRYKM
                     KRQAKDKASQQQMQQDNSSCQQQQSPRRVAVPVLVKDGKPCQAGSNTPTAALQSHQQQ
                     TATAITITTNGLGPHQSHQTNSAGQSPDLVPHSNSPSSLQNQVTSLSHLNSSSSDYGS
                     AMSCSTLLYGRTW"
BASE COUNT      277 a    311 c    265 g    190 t      1 others
ORIGIN


  Query Match             68.0%;  Score 17;  DB 5;  Length 1044;
  Best Local Similarity   100.0%;  Pred. No. 8.6;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaagcgc 20
          |||||||||||||||||
Db    556 TCTGCTGTTTGAAGCGC 540


RESULT  21
AF281080/c
LOCUS       AF281080     1614 bp    mRNA            VRT       01-OCT-2000
DEFINITION  Xenopus laevis homeobox transcription factor Nkx2-1 (Nkx2-1) mRNA,
            complete cds.
ACCESSION   AF281080
VERSION     AF281080.1  GI:10442649
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 1614)
  AUTHORS   Small,E.M., Vokes,S.A., Garriock,R.J., Li,D. and Krieg,P.A.
  TITLE     Developmental expression of the xenopus Nkx2-1 and Nkx2-4 genes
  JOURNAL   Mech. Dev. 96 (2), 259-262 (2000)
  MEDLINE   20417824
REFERENCE   2  (bases 1 to 1614)
  AUTHORS   Small,E.M., Vokes,S.A. and Krieg,P.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2000) Cell Biology and Anatomy, University of
            Arizona Health Sciences Center, 1501 N. Campbell Ave., Tucson, AZ
            85724, USA
FEATURES             Location/Qualifiers
     source          1. .1614
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /note="expressed in ventral forebrain, lungbud primordia
                     and thyroid anlage"
     gene            1. .1614
                     /gene="Nkx2-1"
     CDS             162. .1205
                     /gene="Nkx2-1"
                     /note="thyroid transcription factor; TTF-1; T/ebp; chicken
                     and mouse Nkx2-1 ortholog; contains homeodomain"
                     /codon_start=1
                     /product="homeobox transcription factor Nkx2-1"
                     /protein_id="AAG17405.1"
                     /db_xref="GI:10442650"
                     /translation="MSMSPKHTTPFSVSDILSPLEESYKKVAMEGAGLGAPLTAAAYR
                     QSQVSQASMQQHGMGHNGPVSAAYHMTAAGVPQLSHTTMGGYCNGNLGNLGNMSELPP
                     YQDTMRNSSATGWYGANPDPRFSTISRFMGPSGGMNMGGLGNMGSLGDVGKSMTPLQA
                     TPRRKRRVLFSQAQVYELERRFKQQKYLSAPEREHLASMIHLTPTQVKIWFQNHRYKM
                     KRQAKDKASQQQIQQDNSSCQQQQSPRRVAVPVLVKDGKPCQAGSNTPTAALQSHQQQ
                     TATAITITTNGLGPHQSHQTNSAGQSPDLVPHSNSPSSLQNQVTSLSHLNSSSSDYGS
                     AMSCSTLLYGRTW"
BASE COUNT      475 a    407 c    399 g    333 t
ORIGIN


  Query Match             68.0%;  Score 17;  DB 5;  Length 1614;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaagcgc 20
          |||||||||||||||||
Db    717 TCTGCTGTTTGAAGCGC 701


RESULT  22
AE001220
LOCUS       AE001220    14268 bp    DNA             BCT       16-JUL-1998
DEFINITION  Treponema pallidum section 36 of 87 of the complete genome.
ACCESSION   AE001220 AE000520
```

```
VERSION     AE001220.1  GI:3322705
KEYWORDS    .
SOURCE      Treponema pallidum.
  ORGANISM  Treponema pallidum
            Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
REFERENCE   1  (bases 1 to 14268)
  AUTHORS   Fraser,C.M., Norris,S.J., Weinstock,G.M., White,O., Sutton,G.G.,
            Dodson,R., Gwinn,M., Hickey,E.K., Clayton,R., Ketchum,K.A.,
            Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J.,
            Khalak,H., Richardson,D., Howell,J.K., Chidambaram,M.,
            Utterback,T., McDonald,L., Artiach,P., Bowman,C., Cotton,M.D.,
            Venter,J.C. et,al.
  TITLE     Complete genome sequence of Treponema pallidum, the syphilis
            spirochete
  JOURNAL   Science 281 (5375), 375-388 (1998)
  MEDLINE   98332770
REFERENCE   2  (bases 1 to 14268)
  AUTHORS   Fraser,C.M., Norris,S.J., Weinstock,G.M., White,O., Sutton,G.G.,
            Dodson,R., Gwinn,M., Hickey,E.K., Clayton,R., Ketchum,K.A.,
            Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J.,
            Khalak,H., Richardson,D., Howell,J.K., Chidambaram,M.,
            Utterback,T., McDonald,L., Artiach,P., Bowman,C., Cotton,M.D.,
            Fujii,C., Garland,S., Hatch,B., Horst,K., Roberts,K., Watthey,L.,
            Weidman,J., Smith,H.O. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .14268
                     /organism="Treponema pallidum"
                     /db_xref="taxon:160"
     gene            98. .2149
                     /gene="TP0421"
     CDS             98. .2149
                     /gene="TP0421"
                     /note="similar to GB:AE000783 percent identity: 27.26;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC65409.1"
                     /db_xref="GI:3322706"
                     /translation="MRGGWCLSRRSVGCAGALLLLFAVGAQEHAPPVVRDTALAAGEF
                     RRGVLAYYRGAFNDAIQLFERALSAAPRNPLILEWLGNAYYRSGIEGAALHQWGAARD
                     LGYGGALLRNKIEVVQQRRDFAPDSADALHFSESESFHAVRRGTVLFRRPLSLCALAD
                     GTFWMSAYGSNELLRFDVNGRVIARTRGPVEGFDRPFDVIQTRSGDLLVSEFASDRIC
                     RLTKEGRFLKSFGGKGRGVGQLIGPQFLATDRYDNIYVTDFGNARVAVFAPDGAPLFH
                     FGQKSARFFGFSAPGGIAILDELVYVADALKGAIYVFDTAGNYVRTLVAEGTLKHVES
                     VRAWNGRLLVSLPNEVMVVDVGLASLHTIARLGNAPIRLTAAVPDANGNVLLADYKNE
                     RIEIVSRISELAGGLFVHFERVHADRFPSVTVNVSVHTREGTPVVGLDVSNFFLTEEH
                     RPVSEMRLQGAAYLNDTCSLSILVDRSPATEQEMRLVTRVIGELAQAVGEQGTLSVVS
                     AAQSPVLEGTFTAAQLQAQPLRLRAKSSSQWRFDLGLRLAAGTLVNADYKRAVVFLSL
                     GELQTQDFAQYGVTDLAAYLTNNETSFYVIQLQRTVLPVELGYLVQKTRGAVYSVYAQ
                     EGLGALVAHVRSGAAGTYALSYRSRLPTDFGRAYMPVEAEVRVLSRSGRDETGYFAPL
                     E"
     gene            2133. .3158
                     /gene="TP0422"
     CDS             2133. .3158
                     /gene="TP0422"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="T. pallidum predicted coding region TP0422"
                     /protein_id="AAC65418.1"
                     /db_xref="GI:3322715"
                     /translation="MRRLSDRLRGMTGKGWSAWVRLCVLSTQVGLAAEPALARLAGVW
                     ENATRIVAIAPTADAMRFSLMLVHKPYYAYVYERVAVFSLSSLSDARADVLPPPAAQK
                     AERVDLVLSFEGRVRVQPLVLCRWGDALFCSFFQRVSYGTRSGSVSSEQIPVRTDDPL
                     FGFWIEEGSAEALRAYSSQAPGELNAFFFTEANFYRFRYWRDDALAFAAQRAYFTADD
                     GVTYEIPQYVQRGAAVYSCTTGRSRVVRNFQTGTYEVRTSSDGSKRLMLRRRGAGPGS
                     YAVGAVYPRQGFLDESALALRVDGQFLAIGEPFLRKSAVHDVTAFIDAHNARKRAPRA
                     PLLVPDF"
     gene            3183. .3989
                     /gene="TP0423"
     CDS             3183. .3989
                     /gene="TP0423"
                     /note="similar to GB:AE000783 percent identity: 39.66;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC65410.1"
                     /db_xref="GI:3322707"
                     /translation="MRVFVSWHKQKDKGERERAGALQGACIGTPEHRQVFALVGRSGS
                     GKSFRAQIVMRRFDIAFVIDDGLLIREDKIIAGRSAKQERTLLAAIRVALFENTMHRR
                     VVARVLTYFLCGSQNKKVLILGTSERMVRRIALRVGLPEPGHIIRIEDIASSEEIACA
                     LRTRHVEGTHVIPVPSGEVRKSYPKIFYERIKLLLRREAGAERIGRWAHAIWHEGLKR
                     ACSGAHAHVFEKSIVRPPFSCNLRVVAATEGTQDASPVVVPPHEELALQQ"
     gene            4192. .4890
                     /gene="TP0424"
     CDS             4192. .4890
                     /gene="TP0424"
                     /note="similar to GP:1510115 percent identity: 37.59;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="V-type ATPase, subunit E, putative"
                     /protein_id="AAC65411.1"
                     /db_xref="GI:3322708"
                     /translation="MLGESREEAERIVRAAREEAERIVRAAREEAERIESSSLAALSQ
                     ASRNVLLSFQDSVTRSLRAIISMETAQAYDAGVLRELIPRVVSAWVQAEGDKLELILS
                     PADLRTLEGVFCAALQEQLSAGVELRSDDCLTAGFRIVPAEGGSYYDFSAAAVAQLFS
                     SYVSARVAEVLSLLRRSCDVFLLLPYNAATISSMRRGASFALSGFFGLCATFLGSAGC
                     CGAWKYLVVSRAGA"
     gene            4868. .5128
                     /gene="TP0425"
     CDS             4868. .5128
                     /gene="TP0425"
                     /note="hypothetical protein; identified by Glimmer;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="T. pallidum predicted coding region TP0425"
                     /protein_id="AAC65419.1"
                     /db_xref="GI:3322716"
                     /translation="MCPVPARDSTGSRVLDRWGEFEYALRCALARERADRLRWDTPVE
                     VRQCPVAQESVARAARVAVSLDDPLEAGVFFERRALFLCRGV"
     gene            5270. .7039
                     /gene="TP0426"
     CDS             5270. .7039
                     /gene="TP0426"
                     /note="similar to GB:AE000783 percent identity: 63.40;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="V-type ATPase, subunit A (atpA-1)"
                     /protein_id="AAC65412.1"
                     /db_xref="GI:3322709"
                     /translation="MTQTKGIVSAVNGNMVSVTFEGVVSLNEVGYVHVGNARLKAEII
                     RVRGREAQLQVFEITRGVSVGDRVEFTGDLLSVELGPGLLGQVYDGLQNPLPLLAEKV
                     GYFLERGVYLPALSRTSEWMFTPHVSVGERVVRGDVLGYTPEGALKHRIMVPFHMGDS
                     YEVVFIQTAGTYRVHDVIARVRDAQGHEHELTMAFRWPVKRPVHCYAERLKPTEPLVT
                     SIRTIDTFFPVAKGGTYCIPGPFGAGKTVLQHSTSRNADVDVVVIAACGERAGEVVET
                     LREFPDLTDPRTGRSLMERTVIVCNTSSMPVASREASVYTGVTLAEYYRQMGLDVLLL
                     ADSTSRWAQALREMSGRLEEIPGEEAFPAYLESCIAAFYERAGVVRLRSGEKGSVTIG
                     GTVSPAGGNFEEPVTQATLKVVGAFHGLSRERSDARRYPAVHPLDSWSKYPSVLDARA
                     VAYGRSFLRRGAEVEQMMRVVGEEGTSMEDFLVYLKGSFLDSVYLQQNSFDTVDSAVP
                     VARQKHCYAIVMRVLGSVLAFESKDDARAYFSKLGHMFIDYNCCAWNSEAFVEKEKEI
                     RAFLQGESTKIDSEAEGIIRGME"
     gene            7043. .8335
                     /gene="TP0427"
     CDS             7043. .8335
                     /gene="TP0427"
                     /note="similar to GB:AE000783 percent identity: 72.96;
                     identified by sequence similarity; putative"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="V-type ATPase, subunit B (atpB-1)"
                     /protein_id="AAC65413.1"
                     /db_xref="GI:3322710"
                     /translation="MYKVYEKIESINGSVITVRAKGVHYGELARVKTAFGDSLAEVNK
                     LEGDVVSLQVFAGGRGISTGSEVRFLGRGMLVSFSDYLLGRIFNGSGVPRDGGPALKE
                     GRVEIGGPSVNPAKRVIARRMIRTGIPMIDVFNTLVVSQKLPIFSSSGEPYNELLARI
                     AMQAEVDVIVLGGMGLKHDDYLYFKSALEEAGALSRAVLFVHTASDPTVECLMVPDMC
                     LAVAEQFALKGRDVLVLLTDMTNFADAMKEIAIIQEQVPSNRGYPGDLYSQLASRYEK
                     AVDFDDAGSVTILAVTTMPGDDVTHPVPDNTGYITEGQFYLKNGRIEPFGSLSRLKQN
                     VNGRTRADHRALMDNMIKLYAAYRDALEKRSMGFAMSEWDEKLLTYGGLFEAQLMDLS
                     VNIPLEEAFDTGWKILGRCFTPEETGIRSDLINTYWPA"
     gene            8347. .8967
                     /gene="TP0428"
     CDS             8347. .8967
                     /gene="TP0428"
                     /note="similar to GB:AE000783 percent identity: 45.69;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="V-type ATPase, subunit D (atpD-1)"
                     /protein_id="AAC65414.1"
                     /db_xref="GI:3322711"
                     /translation="MAVRLTKNELKRQKETLKTFRRFLPTLQLKKQQLYAEIRAVEAR
                     ASEVRSRADALRQDVQQWVAVFSESGVFDARLLRVQEVKKTYASIAGVRIPCFLEVHF
                     ACPPYDLYRVPLWVDTAVRKMQEVLSLDLEAHVLDEQALLLGAELRTTTQRVNLFEKV
                     KIPETRACIRKITVYLGDQQVAAVVRGKMSKKNLVDVSRMEQEDTR"
     gene            8964. .10832
                     /gene="TP0429"
     CDS             8964. .10832
                     /gene="TP0429"
                     /note="similar to GB:L77117 SP:Q57675 PID:1590960 percent
                     identity: 25.19; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="V-type ATPase, subunit I (atpI-1)"
                     /protein_id="AAC65415.1"
                     /db_xref="GI:3322712"
                     /translation="MIVPMKKVTLLVLGSEQERSLQALRSFGAVHVQLRECASEQLAE
                     LHALDARCVQAIALVTDAQTKNVTRGEECRVAGQVVEAAEAVEQIVRTHSDRVELAQR
                     IAQCIAHLERCEPWGDFDPADVRALAQRGIHLIPVELSERSYRCLPDELQTLCLARRG
                     GLVRCVLVADKPGLPPSLPADARALELPDVSPADLFVRLRQLREECATLTQRLLAYSE
                     YQGAIRALRQKIAADIEFERVHLSMVSVDVSQWRETGETLRIAHVSGYLPVSRVRAFS
                     ECARKEAWAYCCVDPMPEDPVPTQLRNNRWVNLISPLMNFLGTVPGYWEVDISGFFLL
                     FFGVFFSIIFADAGYGAVLTLVSLGGIVLSKRKHAVVSPAWCLGLYLGTLTMVWGALV
                     CNWFGVPVQYVPASLARIAVWEISGFADAAQRNKNQMHVCFFLGLLHLCLGHLIVVRR
                     TFRSLRVLAEFGSLLMLGGMYVVVLNLIVDKERYPLTGMIVGSIIAGFVLNFIFVNYR
                     VSVRQSVADSMKNVINALLGIVNVFADVMSYIRLWAVGLAGGAISATVNEMTHPLFAN
                     FLAFLGIVLLLFGHGLNYVMSILSVIVHGVRLNTLEFSNHVGLMWTG1RYTPFRER"

  Query Match             68.0%;  Score 17;  DB 1;  Length 14268;
  Best Local Similarity   100.0%;  Pred. No. 9.3;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gctgtttgaagcgcagc 23
           |||||||||||||||||
Db    8188 GCTGTTTGAAGCGCAGC 8204


RESULT  23
AX120335/c
LOCUS       AX120335      801 bp    DNA             PAT       11-MAY-2001
DEFINITION  Sequence 251 from Patent EP1108790.
ACCESSION   AX120335
VERSION     AX120335.1  GI:14037050
KEYWORDS    .
SOURCE      Corynebacterium glutamicum.
  ORGANISM  Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
            Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
  TITLE     Novel polynucleotides
  JOURNAL   Patent: EP 1108790-A 251 20-JUN-2001;
            KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Corynebacterium glutamicum"
                     /db_xref="taxon:1718"
BASE COUNT      184 a    221 c    215 g    181 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 6;  Length 801;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 ctgtttgaagcgcagc 23
           ||||||||||||||||
Db     784 CTGTTTGAAGCGCAGC 769


RESULT  24
AX064067/c
LOCUS       AX064067      924 bp    DNA             PAT       24-JAN-2001
DEFINITION  Sequence 349 from Patent WO0100843.
ACCESSION   AX064067
VERSION     AX064067.1  GI:12541779
KEYWORDS    .
SOURCE      Corynebacterium glutamicum.
  ORGANISM  Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   Pompejus,M., Kroeger,B., Schroeder,H., Zelder,O. and Haberhauer,G.
  TITLE     corynebacterium glutamicum genes encoding metabolic pathway
            proteins
  JOURNAL   Patent: WO 0100843-A 349 04-JAN-2001;
            BASF AKTIENGESELLSCHAFT (DE)
FEATURES             Location/Qualifiers
     source          1. .924
                     /organism="Corynebacterium glutamicum"
                     /db_xref="taxon:1718"
     CDS             101. .904
                     /note="RXA01757"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC25274.1"
                     /db_xref="GI:12541780"
                     /translation="MPTASPIYDVVVVGAGISGLIATQLLDRAGLNIKCFEACSRVGG
                     RAVSVQQSDLFLDLGATWFWLNEPLVQQLVNNLGLGTFPQAIEGDALFETLVDAPSRL
                     RGNPIDAASGRFQAGASSLALGLAAQLKPGVLELGDPVHSLSEEDGEIVVKSSKQIVR
                     AKHVIIAVPPALAAELIGFTLDLPADVRKAAHPQHIAVMNWAKEKYTLPTQAASAGGF
                     GHELFQQPLGHGRIHWASTEVATEFGGHLEGAVRAGIQAALQTGFNLKS"
BASE COUNT      214 a    252 c    239 g    219 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 6;  Length 924;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 ctgtttgaagcgcagc 23
           ||||||||||||||||
Db     884 CTGTTTGAAGCGCAGC 869


RESULT  25
YLI012084
LOCUS       YLI012084    6887 bp    DNA             PLN       23-SEP-1999
DEFINITION  Yarrowia lipolytica pex10 gene for peroxin 10.
```

```
ACCESSION   AJ012084
VERSION     AJ012084.1  GI:5921565
KEYWORDS    peroxin 10; pex10 gene.
SOURCE      Yarrowia lipolytica.
  ORGANISM  Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1  (bases 1 to 6887)
  AUTHORS   Le Dall,M.T., Gaillardin,C. and Nicaud,J.M.J.
  TITLE     PEX10 is essential for peroxisome biogenesis in Yarrowia lipolytica
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6887)
  AUTHORS   Nicaud,J.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-1999) Nicaud J.M., Laboratoire de Genetique des
            Microorganismes, INRA-CNRS, BP 01, Tiverval-Grignon, F78850, FRANCE
FEATURES             Location/Qualifiers
     source          1. .6887
                     /organism="Yarrowia lipolytica"
                     /strain="W29"
                     /db_xref="taxon:4952"
     gene            complement(2293. .3357)
                     /gene="pex10"
     CDS             complement(2293. .3357)
                     /gene="pex10"
                     /function="essential for peroxisome biogenesis"
                     /codon_start=1
                     /evidence=experimental
                     /product="peroxin 10"
                     /protein_id="CAB56484.1"
                     /db_xref="GI:5921566"
                     /translation="MSDNTTIKKPIRPKPIRTERLPYAGAAEIIRANQKDHYFESVLE
                     QHLVTFLQKWKGVRFIHQYKEELETASKFAYLGLCTLVGSKTLGEEYTNLMYTIRDRT
                     ALPGVVRRFGYVLSNTLFPYLFVRYMGKLRAKLMREYPHLVEYDEDEPVPSPETWKER
                     VIKTFVNKFDKFTALEGFTAIHLAIFYVYGSYYQLSKRIWGMRYVFGHRLDKNEPRIG
                     YEMLGLLIFARFATSFVQTGREYLGALLEKSVEKEAGEKEDEKEAVVPKKKSSIPFIE
                     DTEGETEDKIDLEDPRQLKFIPEASRACTLCLSYISAPACTPCGHFFCWDCISEWVRE
                     KPECPLCRQGVREQNLLPIR"
BASE COUNT     1793 a   1841 c   1485 g   1768 t
ORIGIN
```

Query Match 64.0%; Score 16; DB 8; Length 6887;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      2 gatctgctgtttgaag 17
          ||||||||||||||||
Db   4051 GATCTGCTGTTTGAAG 4066
```

```
RESULT  26
AE002054
LOCUS       AE002054    11823 bp    DNA             BCT       22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 191 of 229 of the complete
            chromosome 1.
ACCESSION   AE002054 AE000513
VERSION     AE002054.1  GI:6460010
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
REFERENCE   1  (bases 1 to 11823)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et,al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 11823)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .11823
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="1"
     gene            103. .390
                     /gene="DR2207"
     CDS             103. .390
                     /gene="DR2207"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF11762.1"
                     /db_xref="GI:6460017"
                     /translation="MDMSHDTFEARALQLAYEMTAHNTYEGKVGQGHNEARAADIRRL
                     FDILCGHDPEDHAHESHKKAGRKASEKAAEKRRKKFADLPPVAEAEALGRS"
     gene            complement(449. .829)
                     /gene="DR2208"
     CDS             complement(449. .829)
                     /gene="DR2208"
                     /note="similar to SP:Q60003 percent identity: 48.36;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="lactoylglutathione lyase, putative"
                     /protein_id="AAF11756.1"
                     /db_xref="GI:6460011"
                     /translation="MRALETCLYADDLSAAEAFYSGVLGLTLHSKVAGRHLFYRLDGS
                     MLLIFNPEASAKPGDVPSHQGKPGGHACLAIPREQTAEWETRLRGHGLDVTRYAWGER
                     GESLYFDDPAGNVLELAPASIWGL"
     gene            complement(993. .1904)
                     /gene="DR2209"
     CDS             complement(993. .1904)
                     /gene="DR2209"
                     /note="similar to PID:2338556 percent identity: 50.20;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="SudD-related protein"
                     /protein_id="AAF11757.1"
                     /db_xref="GI:6460012"
                     /translation="MSAHTDWDDAGWNDSGSDHSDWDADPGQKLRRKTRKPQGRRKLA
                     DLTGEETDDIADDTIRRLLDLGHLTEVVCELKSGKEATAYLAETVRGPALLKIYRDFA
                     ARSFKNDAIYREGQVVLDQRARRAMESRSRKGLEMLQFDWVMAEYAHLWQLWQAGLSV
                     PEPLVGPHVKTYAETVPAVLMRLIGTAESPAPRLSDVRLTPEEARRAWQQSVQGMADL
                     LRLGYAHGDYSTYNLLWWENTVTIIDFPQLTTRANPNFKELLRRDAQSLATSFRKHGL
                     QETAEGTLREVQRRALGPAPEPRVVLP"
     gene            2390. .2773
                     /gene="DR2210"
     CDS             2390. .2773
                     /gene="DR2210"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF11763.1"
                     /db_xref="GI:6460018"
                     /translation="MKGYCSGLLLLALLALSPLQSRAQAQTSCNYPVGTFDQVYCVNK
                     VTLQADTEMNTAYQKLLRKLPVSAQQTLRQTQRAWMAKRDNDCVDGYTVYVNCTHDMT
                     VSRLNFLNDRLRECNAAGCQPSKLR"
     gene            2813. .3505
```

```
                /gene="DR2211"
     CDS        2813. .3505
                /gene="DR2211"
                /note="similar to GB:L10328 SP:P19396 GB:M24503 PID:147866
                PID:290501 percent identity: 69.30; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="rRNA methylase SpoU"
                /protein_id="AAF11758.1"
                /db_xref="GI:6460013"
                /translation="MAPTPERYAKIRRVLSKRQPTLTVLMDEVNKPHNLSAIIRTCDA
                VGVPEAHAVPPKHGTLASFEGHTYEATSGSAHKWVKVHPHADAVSAVRELQAQGFQVL
                ATHLSQRSVDYREPDYTKPTCVLLGAEKWGVSDEAAEAADANIVIPMFGMVQSLNVSV
                AAATILFEAQRQRLAAGMYEEPQLSPEELHRLSFEWGYPDLAPGYRERGEAYPALGEE
                GELLREPVSRPE"
     gene       3516. .4277
                /gene="DR2212"
     CDS        3516. .4277
                /gene="DR2212"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF11764.1"
                /db_xref="GI:6460019"
                /translation="MEAMTDALRASAIEKGQKLIGLYRRGVGGERANAGRLLHAHLRA
                HDLTLFDLDGSLPVTQDVAALEGWRESAALLASVLGGALPQDAQDEALTRLVDADDLT
                DAEVARLLPLVDLETLIDVRADGWAYAAGSADGEAEQADSYRQAGRRVQPADILAYPG
                SLAERLREATLRAHFVQTHPERLIRAAPGVPQRFLLGVLASVTGRPAVPDPEGARAHL
                DAGQLARVRALLANHAAQAEAIALDAAEAFGRTLR"
     gene       4297. .5313
                /gene="DR2213"
     CDS        4297. .5313
                /gene="DR2213"
                /note="similar to GP:3334802 percent identity: 44.00;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF11759.1"
                /db_xref="GI:6460014"
                /translation="MSHPRFHPRVLFTKNFSVTSAQIRALEDSPYEVWASHSDPHHGM
                LAAAPHTLTEPRGLLGDDYVAWLLGTCAEHGFDLVLPGKERERLAAYRDTFAAQGTQL
                VVPADEATQRHLERKDEFLRGWDTDILPIPRWTTFQSVAEFDAACAQLRQDGVRLCMK
                PARGIYASGFRVLTERPDLAPFLRGELYQLSVAAARELLAAGELPVMLLMHTLEGAER
                SIDCVAWEGRLIRAVVRRKGAHGQVIEDRPDLVEAAARIAQHYALSGIFNFQTKDQAG
                AAHMLEINARASGGLRYSMAAGVNFARLLLDAATGRLDWDALPAVETGLTVAEEKGVR
                AWRA"
     gene       5301. .6416
                /gene="DR2214"
     CDS        5301. .6416
                /gene="DR2214"
                /note="similar to PID:1695869 percent identity: 61.88;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF11760.1"
                /db_xref="GI:6460015"
                /translation="MAGMKGGVSAAQELTVSLPSGTLTLQVERTLYPPAELLDYAVRR
                NPKRGFLFVSRVLGKHIPIAPATAARTWRDLAASLPPLTAPHFIGLAETATALGEGVY
                RAWRQQTGQAATFQHTTRYVTDRPVLLRFDEPHSHAPAHLLYDPGPAARAARELVLVD
                DEVSTGTTLENLALEWLARHPHVTRVVLVSLTDWCPRREALCAALGVPVDFVSLLRGR
                FAFTPAPDWCPPALPAVTGSGADKTALLPARAARFGDDVPLTFPELGLSPTDRVLVLG
                TGEFQFPAYALAAWLAPQVAECRWNATTRSPILPGRAIRHALSFTDNVGDQMPNYLYN
                VDPAEYTRIFVAYEGACVPDPALLRQLGPHAQAVRLL"
     gene       6413. .7195
                /gene="DR2215"
     CDS        6413. .7195
                /gene="DR2215"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF11765.1"
                /db_xref="GI:6460020"
                /translation="MSHGEAPLILAFADLDDTLFQTLRKLPGHGRAALTPATVNTRGE
                AHSYCTPAQTALLDLLTAGAVTVIPVTGRDPAAMRRVTLPFRSWRVLDHGLTILTPAG
                ELDSEWAAQVQAHLTPLQGTLTALTADMRAHADALGCRLTVHRAGDLPFMTVIKHPDA
                DPEALAALQTRWEAAVDDTDGTHPLHVIANANNVSLLPRAPGKAQAVRWLRERHFPGA
                VLTLGLGDSVSDLPFLRECDFAVTPAQGQLLRATAGVRLPQR"
     gene       7253. .8008
                /gene="DR2216"
     CDS        7253. .8008
                /gene="DR2216"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF11766.1"
                /db_xref="GI:6460021"
                /translation="MQLVQEQRGHGLTLRIHALTHVTRFSQGADGLRETREMDGRFTV
                EAELSGGGVLLEPGAFQYSVGQVQASVEQQQQGGFLRRAIATAGTGESAFATHLSGHG
                RVWTEPSRRHFIIAEATGESGDDMILDDRAFYLAQDTMQLGTHTHTGFSGALSGGGMR
                QPKLSGAGLFVVESPVPVSEVQVIELDGRDTVTVDGDLMLMYSASLRVELRPLVRGLR
                NAARSGEGLVFVLSGKGTVFLTPTHQALSGASV"
     gene       9265. .10326
                /gene="DR2218"
     CDS        9265. .10326
                /gene="DR2218"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF11767.1"
                /db_xref="GI:6460022"

  Query Match             64.0%;  Score 16;  DB 1;  Length 11823;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 ctgtttgaagcgcagc 23
           ||||||||||||||||
Db    3308 CTGTTTGAAGCGCAGC 3323


RESULT  27
AE004026
LOCUS       AE004026    13856 bp    DNA             BCT       15-JUN-2001
DEFINITION  Xylella fastidiosa 9a5c, section 172 of 229 of the complete genome.
ACCESSION   AE004026 AE003849
VERSION     AE004026.1  GI:9107228
KEYWORDS    .
SOURCE      Xylella fastidiosa 9a5c.
  ORGANISM  Xylella fastidiosa 9a5c
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xylella.
REFERENCE   1  (bases 1 to 13856)
  AUTHORS   Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
            Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrer,H.,
            Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
            Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
            Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
            Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
            Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
            Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
            and Marino,C.L.
  TITLE     The genome sequence of the plant pathogen Xylella fastidiosa. The
            Xylella fastidiosa Consortium of the Organization for Nucleotide
            Sequencing and Analysis
  JOURNAL   Nature 406 (6792), 151-157 (2000)
  MEDLINE   20365717
   PUBMED   10910347
```

```
REFERENCE   2  (bases 1 to 13856)
  AUTHORS   Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
            Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
            Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
            Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
            Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
            Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
            Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
            Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
            Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
            Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
            Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
            Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
            Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
            Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
            Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
            Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
            Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
            Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
            Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
            Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
            de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
            Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
            Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
            Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
            Tsuhako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
            Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
            Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
            13083-970, Brazil
FEATURES             Location/Qualifiers
     source          1. .13856
                     /organism="Xylella fastidiosa 9a5c"
                     /db_xref="taxon:160492"
                     /clone="9a5c"
     gene            complement(195. .803)
                     /gene="XF2101"
     CDS             complement(195. .803)
                     /gene="XF2101"
                     /note="similar to GI|1652989 (percent identity: 30 %/query
                     alignment coverage:  87.6 %/subject alignment coverage:
                     67.6 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF84900.1"
                     /db_xref="GI:9107229"
                     /translation="MSRCITGYRWWQWICVVVVLVFFWLVAVAGWIVWIGQRDQAASS
                     DVIIVLGAAAYDAKPSPVFEERIRHGLYLYQHGYAPTLIFTGGFGGSSARFAESQVAR
                     RYALRYDVPPRAILIETVSRTTRQNLIQARQLMRVYGLRRAIVVSDPLHMARALRLCR
                     ELDIDALASATPSTRFRSLHTRWRFLLQEVYFFHRDLVIQAN"
     gene            complement(800. .1042)
                     /gene="XF2102"
     CDS             complement(800. .1042)
                     /gene="XF2102"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84901.1"
                     /db_xref="GI:9107230"
                     /translation="MARVIMFYSSFALSVKSCIGVLTGMLLLHAYVFGCGLGGEFRAL
                     CCDVAWLCHVDISARLSGCGCSAGFFICGVSVLVVS"
     gene            1397. .1732
                     /gene="XF2103"
     CDS             1397. .1732
                     /gene="XF2103"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84902.1"
                     /db_xref="GI:9107231"
                     /translation="MKARLFLPILIMSAISGCATYTYSNGYYHGTPTVEYHYPPNYYY
                     PHGYGMHDLFNYPIPPIYPRDFYRTHIHLPYWQRPHHDPSRPLPPKQPIKQEQPNHIE
                     MGLRDKQPI"
     gene            1914. .2024
                     /gene="XF2104"
     CDS             1914. .2024
                     /gene="XF2104"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84903.1"
                     /db_xref="GI:9107232"
                     /translation="MNIADQDLMLLCSELELMKICPHCVDTLKHQRLAAR"
     gene            complement(2067. .2225)
                     /gene="XF2105"
     CDS             complement(2067. .2225)
                     /gene="XF2105"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84904.1"
                     /db_xref="GI:9107233"
                     /translation="MAAWSREEAWMEFYVFCVIFRNALFRVGTFYAVDVVAPAAALRL
                     QAQRSGLE"
     gene            2271. .4160
                     /gene="XF2106"
     CDS             2271. .4160
                     /gene="XF2106"
                     /note="similar to SP|P25756 (percent identity: 65 %/query
                     alignment coverage:  98.6 %/subject alignment coverage:
                     98.4 %); identified by sequence similarity; putative; ORF
                     located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="glucose inhibited division protein A"
                     /protein_id="AAF84905.1"
                     /db_xref="GI:9107234"
                     /translation="MTDSFYGYDVIVIGAGHAGSEAALAAARTGAHTLLLTHNIETIG
                     AMSCNPAIGGIGKGHLVKEIDALGGAMAHAADAAGIQWRTLNASKGPAVRATRCQADR
                     ALYRAAIRQMIEGQARLNIFQAEVDDLLFEGDTVCGAITHTGLHFKAPAVVLTAGTFL
                     AGKIHIGPTQYAAGRMGDPPATMLAARLRERLFTVARLKTGTPPRIDGRTLNYTAMQE
                     QPGDAPRPTMSFIGNSASHPPQVSCWITQTTERTHEIIRAALHRSPLYSGQIEGAGPR
                     YCPSIEDKVVRFAEKNSHQIFVEPEGLNVIDIYPNGISTSLPFDVQLELVRSIRGFEQ
                     AHITRPGYAIEYDFFDPRGLKASLETKAIAGLFFAGQINGTTGYEEAAAQGLLAGLNA
                     ARHVRGLSSWTPRRDQAYLGVLVDDLITHGTNEPYRMFTSRAEYRLQLREDNADARLT
                     AIGRDLGLIDDARWAHFNAKQEAVARECERLSAFWATPGNALGREVKETLGVTLSRET
                     NIIDLMKRPELDYAALMRVPSLGPGVDDAQVAEQVEISVKYAGYLNRQSEEITRQQRH
                     EATAIPLEFDYAAVRGLSTEVLQKLQHIQPQTVGQAQRIPGMTPAAISLLLVHLERMR
                     RNRVA"
     gene            complement(4348. .4560)
                     /gene="XF2107"
     CDS             complement(4348. .4560)
                     /gene="XF2107"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84906.1"
                     /db_xref="GI:9107235"
                     /translation="MFILLMCVHTHPLIDELVEAWRAGPVIRSMYNAIKQDGSSGITE
                     RLPVRWFSWGRASKIDATAAAILASV"
     tRNA            4928. .5001
                     /gene="tRNA-Cys"
                     /note="found by tRNAscan"
```

```
                     /product="tRNA-Cys"
     gene            4928. .5001
                     /gene="tRNA-Cys"
     gene            5207. .5392
                     /gene="XF2108"
     CDS             5207. .5392
                     /gene="XF2108"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84907.1"
                     /db_xref="GI:9107236"
                     /translation="MQADGTTIIGLDISTKRWATAFLSGQNPLSIEDKSDFVLGAALR
                     NSAKNDTTQRLAPIKTA"
     gene            complement(5566. .5892)
                     /gene="XF2109"
     CDS             complement(5566. .5892)
                     /gene="XF2109"
                     /note="hypothetical protein; identified by sequence
                     similarity; putative; ORF located using Glimmer/RBSfinder"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF84908.1"
                     /db_xref="GI:9107237"

  Query Match             64.0%;  Score 16;  DB 1;  Length 13856;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 gatctgctgtttgaag 17
           ||||||||||||||||
Db    2652 GATCTGCTGTTTGAAG 2667


RESULT  28
CBRG01D9/c
LOCUS       CBRG01D9    40031 bp    DNA             INV       02-MAY-1996
DEFINITION  Caenorhabditis briggsae cosmid G01D9.
ACCESSION   U56248
VERSION     U56248.1  GI:1293789
KEYWORDS    .
SOURCE      Caenorhabditis briggsae strain=Gujarat G16.
  ORGANISM  Caenorhabditis briggsae
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 40031)
  AUTHORS   Wu,X and Le,T.T.
  TITLE     The sequence of C. briggsae cosmid G01D9
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 40031)
  AUTHORS   Washington University Genome Sequencing Center.
  TITLE     The C. briggsae Genome Sequencing Project
  JOURNAL   Unpublished (1996)
REFERENCE   3  (bases 1 to 40031)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-1996) Robert Waterston
COMMENT     Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO  63110, USA
                    e-mail: mmarra@watson.wustl.edu
                       NEIGHBORING COSMID INFORMATION:
            The neighboring cosmids, chromosome and orientation of G01D9 are
            unknown.

            NOTES:
            Coding sequences below are predicted from computer analysis, using
            the program Genefinder(P. Green and L. Hillier, ms in preparation).
FEATURES            Location/Qualifiers
     source         1. .40031
                    /organism="Caenorhabditis briggsae"
                    /strain="Gujarat G16"
                    /db_xref="taxon:6238"
                    /clone="G01D9"
     gene           10. .14844
                    /gene="G01D9.5"
     CDS            join(10. .816,859. .1074,1134. .2667,2729. .2930,2975. .13571,
                    13615. .13860,13905. .14078,14121. .14306,14353. .14400,
                    14449. .14558,14609. .14693,14746. .14844)
                    /gene="G01D9.5"
                    /note="Similar to polyketide synthase."
                    /codon_start=1
                    /evidence=not_experimental
                    /protein_id="AAA98699.1"
                    /db_xref="GI:1293790"
                    /translation="MDSLSIVDFVNRLNEKYFPEDEITASDVFDYPTVDELADHIVRK
                    KSVSFPLHPASGETMNGVENPIQNEPEVEQLTQSYILKNSPTGFKLDYLLTWSGLGLQ
                    LSNPSSSNFILALNIDGGQDKELRKMLAEKQNILINFDKLQDSSIETIYMSLLNFVKT
                    LSKSKIKCNFGVSGEFSLGNSIARAFMKTISSEKYPLISFEWDQKIQQVSFTDSNSSI
                    PGNWLITGGLSGIGFEIGKFLGNSGAENVILISRRQPSADVLKEIKTWTAKVHTIAAD
                    VNNKKKLIEELTRVNVDVTGIIHSAGVLKDSKIERQTKDSFNEVFKPKADGVKVLEEI
                    EKHFSYKIEHFIMMSSFTAACGNEGQLNYGVANAFLEHHVQRRRREGKSGCAIQWGNW
                    IDTGMATNPQVRKFLADLGFLGQHNKDALKYLKACIETKPEIVMVANIAWDVILKNRK
                    DLPRDLIINGILPFSTVTIEMNGNEQSPTFSDNFEEVSINVAVEDADEVLELLKEKVS
                    EIVKCAPSKLKSNKNIMDLGMDSRLLVEFLNYINLTFKIKLNLADAYNYSTLEKLAAC
                    IFDKLSNTAEESPMGQDETQEETLVDKRETKSTDFCPYFGVNVFFDKKENLEKAKSEA
                    VKCLVSGKTLPAAGNFAIPVIGTSISDVISKIKAATETQVKTCQRKSKCVLMLTGQGS
                    QYPLMGRQLVENYEVFRATLQSCLAKCDEYLQGEVSLWEILFNTEHYKLLQLTKHMQP
                    IMFSFGIATAQLWLSFGLVPDYYLGHSVGELAAGVLAGIMTLEDGLRLIVERGKAMEN
                    IAGLGALLAVQREIADDVLRKFKVSVAIINSPKQVVFAGTKPELDSALAFVKGQGKQG
                    TFVNQRYPFHSSLIDETHLVALRKCLADIRFRQGATRLVSNVTGQIINTFSEAYIIKH
                    TISAVKFVDCVETLRNDGVAVWIDAGPAAILATFVKRIIVPAELSKHRIVQTCKEKDS
                    DVETLVQSCLELEQSGFEILWPKVYGCDTDTDAPLAHFPVVQSDTIQNHEFEILEGHR
                    LNGKIIVAGAYQMFKMDQLVKSKASGLDLVLKNVKFVKPWYIEDCKAFLIRWNSDMSI
                    ELVVNSVTVCSAKVEMKEGALEIQVIGENEKPFDVHEFYETLYRNGLQYESGFRRMDD
                    SKRSDKRCFSSLKPSPFAWPLIDSAMHSITASVVPRRPDSYFLPVAMGQVTLKQEEDF
                    KLKNLNAQTVITSETDKFIQVNVVLVAGDQIICEVRNMTIVVLKLVPAETPTPTPTIA
                    KEAKKSEIEIVGFDITLPYNQLSTNNENWSFLKTNRVKQPLQTRNAKTEERVALLDAD
                    LRYWDPEYFGIRPSEAGYLDSQQRLLLTSVAKLLDSLSISSLPSKTGVFIGCSTYEFS
                    QIVYAHNQENPRAEWGGGTSNSALAGRIAHWLKLKGPVLTVDTACSSSFHALSSACET
                    IRSGQCECAIVGTVNLVLHEMTSNVLQNAQMTVQDFCKAFDVDANGYKRSEAVCTLLI
                    SKTGKLKSMASIVDWAVGHNGNSSSLFTPNGISQLEIMKKATSNTRDIIGVETHCTGT
                    KLGDPIELSAISKLISTDCSISSSKSNIGHAEGASGLVSLCTLLMSLQTKYKMAQLHL
                    KCPTEVITKGALHCQFVGEDLLADEHSAFLINNFGFTGSNCSLVIKPKQNDPEIHESK
                    DVFYPMLLSAHTIKSLERYISVFREFVASSSASLRDIMLSLLQKKVHVHRQFLIYSFK
                    RKLVSDQFDDERWNPIRQRCMVFVKEGTTVFEKATHFQRVQLPPIVFNNSLHWKLNPF
                    HDEIGQQPCVPLREIFFEKVLVEKAENNEPPMRNVACVGRLTLIPKIEMEEVSCFLNG
                    IIVFHPLSNSIEEFLKLTELWSLVSPDQNSVIIVACFKNGTSHTEWTGVLRSLASEKP
                    ILYKFVSIDHVDHLKKEFNYQDVFESIFYSNGSRYVERLKRLTPTPQKMEEIEKCLIS
                    GGTGGIGKAVLNEIKPKKSVIITRKEPKEEETKSYLSSDVADFISHSKFDHVFHFAGT
                    VKNASHLNIKKDILEEMVLTKFQGAKNLNRCCSKLGKFYFSSSIACILGSFGQSNYTF
                    ANGLVNSFLENLSRNTKIIHWGPWKEVGMLAKSECDDIKKQIEMNGWIPLSNSDALSV
                    FSTDLINSKHQVIVFDGNFEEIVSRQPHLQNLLSEIVLKSTAQKKQFSDFESIFYDVV
                    GIRDMNDKKHIPFMDLGIDSLCMENLRHTLNVHLNLELNVSEIFENATYEKLKNYVAS
                    MLKSKSTEKSSPKKSDNITNFRENDNRVAVIGWSAEFSGAENIEEFWENLLDGVVSTG
                    KENSLLKSPFGFDNKFFNITDEDAKMLDPQVRKFIQHSYLALENSGYVDKRQDLKFAV
                    FAGSEPSDYGRPESQNDAMRKLFVMNMNNYLASYASYCLDLKGETVSVYSACSTTLVA
                    VANAVKTIQSGSVDYALVGTASIAEASECLSGNSDEKKTIFSKSGICRPFDRNSEGIV
                    RGSGVGCLVLKKYSEAKKDNDNIHFVIKDFAISNDGISRASFMAPNPAGQQKCMTEVL
                    EKLTQEEMNRIEYVECHATGTTLGDTIEMNALQTAYCFKKNLQIGSCKANIGHAYAAS
                    GLASLIKCAKMLQTGKVPPQVNFSEFRDSFGTFFNVSTQKSFIQNDSFIAIDSFGIGG
                    TNVHMIIERLGIREEKNDDDVENVILPISANSKASLVEAIKTMEVQLKKSNLNKLSST
                    LLHHRNQMNFRAFIIASKIIGKVQIKSFSKIKTSMKSPKIALFFAPQGIQFANILPHE
                    YVHNFDFKSSLERHCKTAEHLGVSNLLEILYPGNYENEDLIQDTLYSQISLFVQCMAI
                    FECLKAVIKPQLLIGHSVGEYAAAVISQCISEEEMMELLIERSNLISQTQKARMLMVW
                    NYEEELPNSVVISAIIDPITKCIVGPTEAIDELEKTFKLCQVKYRNIKTNHGFHSKML
                    QPVSQKFKNLCLSLPSEPPRIPIISSITGQRIKRFDAEYCRKHLTTAVNLDLVVDNII
                    KSGVDVVVEIGPTGVLSNLLGKKNSKIVVIPTCGTKKHPKASLSDCIGELWCCGIDIR
                    PLFPKLKIDGNAPGYCFDEVQFGGHKKTFPVLKFNFYSEKWSKLNNSLSTTAALSDFQ
                    IYDKNIGSTQTSKILVYHLEGCSDSHQNYFEVEKFLKQCQYETCLFITSSDSASVHII
                    LGLLRCHQIVSRTKIKFVENFENLELKKVVSNAIASDGLYVRVTSEGVFEHGFLQISP
                    NTHSIKTKMRENALVFGASGFVGQSASRLLHSFGTNVIEVSRSSKISSDIRHLKSVQE
```

```
                VLKSLNLKHFDYVLNCVGKESSSK1NKTFEEQEETLTTKTYGNSNILKALQEYNIKVD
                KYITFSSLSSVIPLLGNSDYASANCYVEALKSTFVQKLLVLSLPPLVDSRMYEKSTES
                TKDMLSNLSMNSEELDNALKIALLTETEGTLFVSLEDPAQIARKSLSYHRRDEIGEVP
                ENEQRECLVTSIRNIWKETLGQSLGIDISSNFFSLGGDSLSALQVVWQVQKRTGKTIE
                VNDLFNNPTLQQFTDFVQMSNSQFRPSCHQDDEKIPDFNNIQLTHSQTQMFMLRQVDV
                TSRYNVIFEIVVKYDISFCWNSLKYSILSLIAYQTSYRSVFENSRPPIQKVCSLTESF
                HNLDDRWDLLDAISQELYHSFEIGQTTPLRVRIAEDSDNNQIHIVFNQHHILTDGWSM
                TVLADTISKLYTSFKLKTVFPLKSNQSVAKLAIQQCVENLQEAVDNASKYFENENHTI
                IPYDVSAVKEENHKRFIKRIPTRIWSKIVDMSKKYSTTTHNICLAVFCECVRSFTGQT
                DLMFAFANSGRNSENKDLVGYFMNNALFKTSISEDVEKFEKILTIVSKSLKDCRKCSN
                VPFHTLVEQNRKLNAISIYFNFRQKLDYPSVSIPGATCEIKHLSLNNAFSFSFTIDET
                PTGPLISIDYDSSKYFESTVNLFSETFLKKAKFSFSKAQNNKIIHVRRRADFPSFLFQ
                IDLLTSWKLFSDESPALLLSSNTIMYSELAEIIEKVAKCIQKQYQIERAASVREDDII
                GVYSSKTPISILACSILGASYAPVDPSWPEDRQLYVKSKVNIVFDESHDVNFEVLNLR
                SFKTRTQYAGIYTIFTSGSTGVPKGVFMAEQSVSSFLTSATKQCMFREKVRVLDSVKH
                VFDVSVSNIFGPMLNGGCLISADHSTTITDKLQKCQYAFLPAAVFNGFTSETMRRLRT
                VETLTIGGETVSDAALQCALKMYPKMRIIQIYGPTETCVWSLTNRCRISESENVGANL
                GESMENECYRIVNDSIRGILKVQGISLARGYFSSNPKCFGKNYSTGDVVDSRSGIVKY
                IGRIDNQVKWKGARVDLTELEKELLLCLGLLQVAIVYSNQMLIAFLAGTSTAFQYSDL
                IAKLQNKTHLPDHYVQIDQMPLNSSGKIEKSTLLQAFESVRKSYKREMVPMKNSLEEK
                IVKASGDVLGRKTEITDKFSDVGGNSLTAIQLAHRLTEELKFDVKAHEILQSNNFGDL
                SKTLASSVTKPESKVPNVITKLREVPNSKFSIYLVHAIGGTIYPYYSFLQIFPKNISL
                YGIEFDLKYPCGDLRELAHFYAQEIAAHAGQKQIFVMGHSMGGIMSREIVAELKIWGY
                QIPFVMLFDSWVLRTNELDIENIKQFITYVFSGLPDSEHRIDRAIKLAQLLREYKTSV
                SDTKLYLFKSKQLGDAAFKKAVRVDLNEELSRSMTCNGFDELSLQPIETYLIDGDHES
                CLKVENLKKVKDFILAPFKPYFDQA"
gene            18313. .19760
                /gene="G01D9.4"
CDS             join(18313. .18387,18490. .18569,18615. .18737,18791. .18886,
                19060. .19237,19285. .19395,19444. .19624,19684. .19760)
                /gene="G01D9.4"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA98700.1"
                /db_xref="GI:1293791"
                /translation="MTIFTKTETDTDKKNAEGVKEDNPKVVVCEADEEKHISNCGWVP
                MEGGNFDRPKEKRSSSFGPLSIIRLSPSPLARTPSPSPPDYENAVNARRMRRCTKLTF
                TLIAIFSFVILVSFFLYQQGLISVAYQKGVYHGWNAATFSNRGTTERLVQNVEINAND
                SYEKIDVPKFGSNRPAIFLHDFKQNLTAIVDTVGNRCFVKDLDRTKIRSPRMLIEMLR
                NIDIMAPEVAYAQTRMVRETYNVGDELTKADISTFNSTILSRHCAFRQTYKLRKQEAQ
                TRSRRAAGMGPALSFASMAGDSVQLEEISF"
tRNA            complement(19138. .19236)
                /note="Leu;  codon recognized: CTA"
gene            20621. .21971
                /gene="G01D9.3"
CDS             join(20621. .20772,21464. .21614,21664. .21783,21834. .21971)
                /gene="G01D9.3"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA98701.1"
                /db_xref="GI:1293792"
                /translation="MKIFNLSFLVISYTTKPPCCRDHLGAAACSKLMHQNTRLFAKRC
                NSDAEFRLIQCCSSCNTDGVGMAYDLIARSLVSEHCFDRYGPGFCDRLRSPPPLSNTP
                HNTWSCDGENPQIAFRTCRKSCGYCNFSVVQYTLDNALQACRVQPVAEEKRRWRVKFY
                SSTTTTTSSPSPMEVINSTYQFWNHN"
gene            complement(22269. .22799)
                /gene="G01D9.6"
CDS             complement(join(22269. .22371,22415. .22541,22588. .22680,
                22730. .22799))
                /gene="G01D9.6"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA98702.1"
                /db_xref="GI:1293793"
                /translation="MIPRFLTKVCIISLLIGVPGVTDASTSMYSRRRTPSERVMSIIR
                PRMFQNPEFKIGQTIQKSIQLNPNRALPCCKDESGGSICKNLRRTDIKLFTQKCQTEP
                DFSLVVCCNSCSEAGISYRKRAQKQVNL"
gene            32520. .36916
                /gene="G01D9.2"
CDS             join(32520. .32678,33205. .33505,33554. .33662,33836. .33936,
                33983. .34088,34698. .34775,34824. .35288,35332. .35502,
                35548. .35847,36442. .36637,36689. .36916)
                /gene="G01D9.2"
                /note="Similar to receptor tyrosine kinase."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAA98703.1"
                /db_xref="GI:1293794"
                /translation="MQNGDVLDAQITASSSFDKQSVGPQNARLHSELASGAWCPKPQI
                NSKSYEFLQVTLNDTFLITSVETQGRYGNGTGREFASHYMIDYMRPGSQWIRYLNRSG
                HVYMDGNFDTTTPVIRALDPPIVASRIRIVPSSKNTRTVCMRAEIHGWKHEGVIYYST
                VPDGSRLDTLDFKDNMFENSQMYTESGITRGLGLLTDGYVAQTSPFEKNQMNNSWIGW
                NRDTTDGRVTILFEFEEIHNFTDVVLATFGNRIDGIDVIFSQDGKTFPLFSQISSSER
                QTLNNTSRRYDFRVPLHNRAGRKVRISIKFSSDWMFLTEVHFTSGIFKHTVSQKNINY
                LAANLTLMTEKVPTPESATTQKLSVIGGIIFLLIFVGSIYCISVCLKRRQKNKSVDSN
                VKKDLIITHMGNKPSCHVFPSNGKLSNGHYEVANDILYTRSQKSTLLSVSSKSSSSCR
                VIPPTWNDFNFPPPPPEPNKEEHTYSQPVSPENSSNGSYKSVKKVPALKKYPTSALLIG
                KAIGEGKFTMIKECIIFGGLKCAHKSSKEEDCVHGTRAMGDEISCLLKCGRHPRIVEL
                YGVDESYNLLLEPVEYGCIRNFWLASETSLDTEFLARICKDVFSAMTYLESIRIVHGH
                FTPNNILLDGEFHAKICSPRGPSHHAQLRYSAPESIVNNEFTHKSDAWAVSTTVYEMA
                YQCRQRPYEELTNEQIVDNARALLDHQPHAVRGKFELSVIMFFSACSIDSNSVPL"
gene            38402. .39782
                /gene="G01D9.1"
```

Query Match 64.0%; Score 16; DB 3; Length 40031;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      6 tgctgtttgaagcgca 21
          ||||||||||||||||
Db   7934 TGCTGTTTGAAGCGCA 7919
```

```
RESULT  29
AP002897/c
LOCUS       AP002897   156425 bp    DNA             PLN       08-MAR-2001
DEFINITION  Oryza sativa genomic DNA, chromosome 1, PAC clone:P0686E09.
ACCESSION   AP002897
VERSION     AP002897.2  GI:13365563
KEYWORDS    .
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0686E09.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (sites)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
            clone:P0686E09
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 156425)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2000) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Mar 16, 2001 this sequence version replaced gi:10944300.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'.  A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
```

This sequence of P0686E09 clone has an overlap with P0410E01 (DDBJ:AP002866) clone at the position 1 to 1233 of 5' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES Location/Qualifiers

source 1. .156425
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0686E09"

gene join(4841. .5341,5904. .6008,6478. .6591,7059. .7241, 7450. .7578,7719. .7781,8016. .8111,9084. .9173,9591. .9707, 9900. .10022,10215. .10307,10717. .10878,10992. .11114, 11186. .12178)
/gene="P0686E09.1"

CDS join(4841. .5341,5904. .6008,6478. .6591,7059. .7241, 7450. .7578,7719. .7781,8016. .8111,9084. .9173,9591. .9707, 9900. .10022,10215. .10307,10717. .10878,10992. .11114, 11186. .12178)
/gene="P0686E09.1"
/note="hypothetical protein similar to Arabidopsis thaliana chromosome 1, F26A9.21"
/codon_start=1
/protein_id="BAB39109.1"
/db_xref="GI:13365564"
/translation="MALNMKTLTQALAKASAVIEKTVTTTVQEVTGPRPLQDFELLEQ
AGSGGPGLAWRIYTARPREGAPSAPYPVVSVWVLDKRALAEARARAGLTKAAEDAFLD
LARADAARLVRLRHPGVLHVVQALDETKAAMAMATEPVFASVANALGCLDNVGKVPKE
LKGMVRSEMGLLEVKHGLLQVSETLDFLHNNAHLAHRAISPEQTVFITLNGSWKLGGF
GFALSVDQATGSLTSTQQFHYSDYDVEDTALPLQPSLNYTAPELVRSGDSKVGSACDI
FSFGCLAYHLVARRPLLDCHNNVKMYMNSLTYLTSEAFSNIPADLVVDLQRMLSVDVA
SRPSAMAFTGGSSFFRNDTRLRALRFLDHLLLQERDNMQKTEFLKALSDMWKDFDSRV
LRYKVQVLPPLCAELRNMVMQPMILPMVLTIAESQQDKGDFELSTLPALVPVFTSASG
ETLLLLVKHADLIIHKEHLISHVLPMLVRAYDDTDPRLQEEVLRRTVPLSRQLDVKVQ
LLKQSVLPRVHGLALKTTVAAVCIIHVLVLVRVNALRCLGDLVPSLDKEGILEILQTL
RRCTAVDHTAPTLMCTLGVANAIFKQQCGVEFAAEYVIPLVFPLLTAHQLNVQQFAKY
ILFVKDITRKIEEKRGVTVTDNGNTEVKVSPSLANGIHSEPLSRGVGQTEQIQTAKGA
AWDEDWGPTKKASTPSQSSDSNARMKQPSDPFDFSTQTKQPSALPFDFSTQQPKPSSA
IPQVTPATISPAPSLPSLQSLAPSSGPQTSGSCVPVDIEWPPRKNTSSDFNAPLSVSE
QSKSGGLSSDGLADIDPFADWPPKASSGASILAVGHLSSTSQSISGLSTGNIGFSGNS
GSLGQMKTNQMSWSAKPNTTNLMAVNSTGNYLNQGNSALGFGNPIGGRSTGSLSSSTT
QGPPRLAPPPSASVGRGRGRNQGQSALSRASRTPHSNVSSEQPPILDLL"

gene complement(join(13392. .13540,13940. .14123))
/gene="P0686E09.2"

CDS complement(join(13392. .13540,13940. .14123))
/gene="P0686E09.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB39110.1"
/db_xref="GI:13365565"
/translation="MIHPSIRFAPYLAVFPRSGGYPCRLLAAYPAAPCPPRLQEPPYH
NQCVVAFALQFGGVARAPCIWLDGATVTVEFEGQPASSQLLAKSLVCVHSDVAEMFHM
VHRWTDAP"

gene join(28972. .29115,29117. .29179)
/gene="P0686E09.3"

CDS join(28972. .29115,29117. .29179)
/gene="P0686E09.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB39111.1"
/db_xref="GI:13365566"
/translation="MHAAAAASARTACSAARSTSLERWVGAHGHGHAGVVTIDDGMQR
RAIPYHRIQSAMMRARAGRGREAG"

gene complement(join(29740. .29837,30219. .30380,32426. .32546))
/gene="P0686E09.4"

CDS complement(join(29740. .29837,30219. .30380,32426. .32546))
/gene="P0686E09.4"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB39112.1"
/db_xref="GI:13365567"
/translation="MFGAISGLSPREVFMCVLNLLERVLDGHSRTQPGAIHLQNVLQR
KRGHDLSTENALAAHKGCPFLAVLFTGLQAMTQASAPAPTSSREAQALKHRIPDTECV
GQGTEQRTDAKLLKASRRCDATRT"

gene join(33917. .34076,35467. .35816)
/gene="P0686E09.5"

CDS join(33917. .34076,35467. .35816)
/gene="P0686E09.5"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB39113.1"
/db_xref="GI:13365568"
/translation="MRWTCLAGSAGVLPGRATHVLLTLSRGERQFTILPFALLFPCFL
GCMYGLLVAVLYKDNIETQEAVRKECVPVCVVGKNMKAKSTVFVVGAERSFWRPLEIY
GVLSLLIPSLSVFGKQILWFLGYACVCSVREKAGKLRDRAPCFPLSLSLMNFDGLGKC
NMHPFLKIS"

gene join(37532. .38415,38988. .39218,39280. .39471,39552. .39890, 40001. .40207,40381. .40515,40601. .40765)
/gene="P0686E09.6"

CDS join(37532. .38415,39005. .39206,38988. .39218,39280. .39471, 39552. .39890,40001. .40207,40381. .40515,40601. .40765)
/note="hypothetical protein similar to Oryza sativa chromosome 1, P0665D10.16"
/codon_start=1
/protein_id="BAB39114.1"
/db_xref="GI:13365569"
/translation="MSEADRIRVRAAALALDGGGGGAVRDKPDAKADVFADLGSPVSP
LRARASVATSSSSSSGSAKSPAPSNAGALALAGGRSHSGELTAESTPPRLPGHRRCGS
GPLIFSGGSSGGSGGGGGDRGSTASSPMTNALPAGNICPSGRVPVAAAAPPPPRSRPD
VLGSGTGNYGHGSIMRGGGGMAPARSSIDSSSFLGHAPRSPATFPAASSASSGSLQDV
TRLGNEWYKKGKHAEALRHYDRAVALCPESAACRGNRAAALAGLGRLADALRDCEEAV
RLDPANGRAHSRLAGLCLRLGMISKARRHLTQAGHLHQSDPSEWEKLQEVEMHQGRSI
DARKVGDWKSALREADAAIAAGADSSRLTLHCLRLGMISKARRHLTQAGHLHQSDPSE
WEKLQEVEMHQGRSIDARKVGDWKSALREADAAIAAGADSSRLVMSSLLAIRSEALLR
LHKLEEADSTLASLLKLDSVLLYRMGANPSGMLAESYVSIVRAQVDMALGRCDRFDAA
VEAADNARFIDPGNAEVGMILNNVKLVAKARAQGNELYKAAKFSDASIAYSEGLKYEP
SNPVLYCNRAACWGKLERWEKAVDDCNEALRIQPNYTKALLRRASSYAKVERWADCVR
DYEVLHKELPADTEVAEALFHAQVALKTTRGEDVSNMKFGGEVEMVTSVEQLRAAIGS
PGVIEQFAISSGVSVVYFMSIMNQQCTLITPSVNSLCSECPSLNFLKVQVNVEDSPMV
AKAENVRIVPTFKIYKDGVKVKEMICPSLHVLRYSVRHYAVSSS"

gene complement(join(47016. .47264,47291. .47363,48616. .48748, 48784. .48832,49178. .49324,49569. .49684,50921. .50967, 51070. .51107,51434. .51652,51763. .51861,51888. .51986, 52059. .52112,52817. .52933,53291. .53532,54642. .54895, 55164. .55345))
/gene="P0686E09.7"

CDS complement(join(47016. .47264,47291. .47363,48616. .48748, 48784. .48832,49178. .49324,49569. .49684,50921. .50967, 51070. .51107,51434. .51652,51763. .51861,51888. .51986, 52059. .52112,52817. .52933,53291. .53532,54642. .54895, 55164. .55345))
/gene="P0686E09.7"
/note="hypothetical protein similar to Arabidopsis thaliana chromosome 3, MBK21.5"
/codon_start=1
/protein_id="BAB39115.1"
/db_xref="GI:13365570"
/translation="MSDREVSQTRENRTWDMGYARENKAWDGLLAVKVEGARESKRCA
LGQARHVHILGLRVQGSASRQPPSAKTLAASAASGGASMSDPFYPHGHGGAAGGEGAA
AAGYSSYEVDLIAARYGGRPLANPSSAAADLDARLAGARRSMGVLYHQPIMGSHSTVE
QIEALYSSNTMTKRPRLESSLPIYPQRPGEKDCAFYMMTRTCKFGGSCKFDHPQWVPE
GGIPNWKEHLLFEALDNTLILYKGYIGNSCQHHVFWYLTALGVQPRFAANVEESYPEQ
EGEPDCPFFMKTGKCKFGSKCKFNHPKEKVNALASGNTNDTTNMSSSQPLRSQPVCSF
MKLQKHLIADSSILPFYAKTGKCKFRAMCKFNHPKDIEIPSSQNEPESAVTVEGETDI
GSAADSVSAKMQTPVAAAQEFNSKGLPMRPGEVDCPFYMKMGSCKFGSTCRFNHPDRL
VLNFPLPLGQTILPTPESMLLNSSANFMQGFDFHAAHMPFYMKTGFCKFADRCKFHHP
IDRSAPDPSANWEPAEESVQLTLAGLPRREDAVVCAFYMKTGVCKFGGYRQGVEFREL
RRETNILTQGRIKSMPYSFLAYELNMKPHYLGSFLYRRRPLSELESPSKAKAAAIFSA
QPVHAYGHRLRLAPQVAPLHPTASGREPITRGSLAIAWSPRRAAGGGDAVRLPGCLPV
GCAAWPARSAAASGDHPLCCAGY"

gene join(57612. .57759,58207. .58289,58617. .58775,59266. .59373, 59488. .59548,59653. .59918,60187. .60219)
/gene="P0686E09.8"

CDS join(57612. .57759,58207. .58289,58617. .58775,59266. .59373,

```
                     59488. .59548,59653. .59918,60187. .60219)
                     /gene="P0686E09.8"
                     /note="contains ESTs D40230(S2054),D23027(C2064)
                     hypothetical protein
                     similar to Arabidopsis thaliana chromosome 1, F19K6.5"
                     /codon_start=1
                     /protein_id="BAB39116.1"
                     /db_xref="GI:13365571"
                     /translation="MVGWRAAGGARAVLRRLSAAAEAAAKQDGRVFAASYSGSSGGVN
                     APFGLGQYANLLRAQAFASRGVALNFHQLIRNAGISTTRNLLAADDAMVPVSSPLTPP
                     LGDGEQTDKKGAIVKRLKVQAIKKDIKQSPKKVNLVAKLVRGMRVEDALLQLQVTVKR
                     AAKTVYHSARANAAHNHGLDPDKLIVEEAFVGKGLYLKRLSYHAKGRCGVMVRPRCRL
                     TVVVREATAEEEAKIAKLRVSNYKKLTRKEKQLMPHRLIEVSPRWARKRKEEAGAAAL

  Query Match             64.0%;  Score 16;  DB 8;  Length 156425;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 atctgctgtttgaagc 18
            ||||||||||||||||
Db   52932 ATCTGCTGTTTGAAGC 52917


RESULT  30
AC016697/c
LOCUS       AC016697   163489 bp    DNA             PRI       25-MAY-2001
DEFINITION  Homo sapiens clone RP11-140C4, complete sequence.
ACCESSION   AC016697
VERSION     AC016697.8  GI:14196409
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163489)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163489)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 163489)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On May 25, 2001 this sequence version replaced gi:13431103.
            Center project name: H_NH0140C04.
FEATURES             Location/Qualifiers
     source          1. .163489
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-140C4"
BASE COUNT    46776 a   33045 c   34796 g   48872 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 9;  Length 163489;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gctgtttgaagcgcag 22
            ||||||||||||||||
Db   40201 GCTGTTTGAAGCGCAG 40186


RESULT  31
AC023914/c
LOCUS       AC023914   168177 bp    DNA             HTG       20-SEP-2000
DEFINITION  Homo sapiens clone RP11-44K20, WORKING DRAFT SEQUENCE, 22 unordered
            pieces.
ACCESSION   AC023914
VERSION     AC023914.3  GI:10198362
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168177)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-44K20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168177)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 20, 2000 this sequence version replaced gi:9102875.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1087
                Center clone name: 44_K_20
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 157528 bases at least Q40
                Consensus quality: 162841 bases at least Q30
                Consensus quality: 164870 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 166077; sum-of-contigs
                Quality coverage: 4.3 in Q20 bases; agarose-fp
                Quality coverage: 4.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1263: contig of 1263 bp in length
            *    1264  1363: gap of      100 bp
            *    1364     2399: contig of 1036 bp in length
```

```
            *     2400 2499: gap of      100 bp
            *     2500    3713: contig of 1214 bp in length
            *     3714 3813: gap of      100 bp
            *     3814    27125: contig of 23312 bp in length
            *    27126 27225: gap of      100 bp
            *    27226    28854: contig of 1629 bp in length
            *    28855 28954: gap of      100 bp
            *    28955    31625: contig of 2671 bp in length
            *    31626 31725: gap of      100 bp
            *    31726    34084: contig of 2359 bp in length
            *    34085 34184: gap of      100 bp
            *    34185    36582: contig of 2398 bp in length
            *    36583 36682: gap of      100 bp
            *    36683    39629: contig of 2947 bp in length
            *    39630 39729: gap of      100 bp
            *    39730    43900: contig of 4171 bp in length
            *    43901 44000: gap of      100 bp
            *    44001    46837: contig of 2837 bp in length
            *    46838 46937: gap of      100 bp
            *    46938    52599: contig of 5662 bp in length
            *    52600 52699: gap of      100 bp
            *    52700    60235: contig of 7536 bp in length
            *    60236 60335: gap of      100 bp
            *    60336    67600: contig of 7265 bp in length
            *    67601 67700: gap of      100 bp
            *    67701    76219: contig of 8519 bp in length
            *    76220 76319: gap of      100 bp
            *    76320    85003: contig of 8684 bp in length
            *    85004 85103: gap of      100 bp
            *    85104    95875: contig of 10772 bp in length
            *    95876 95975: gap of      100 bp
            *    95976   108080: contig of 12105 bp in length
            *   108081 108180: gap of      100 bp
            *   108181   126996: contig of 18816 bp in length
            *   126997 127096: gap of      100 bp
            *   127097   145024: contig of 17928 bp in length
            *   145025 145124: gap of      100 bp
            *   145125   166554: contig of 21430 bp in length
            *   166555 166654: gap of      100 bp
            *   166655   168177: contig of 1523 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168177
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-44K20"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1263
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1364. .2399
                     /note="assembly_fragment"
     misc_feature    2500. .3713
                     /note="assembly_fragment"
     misc_feature    3814. .27125
                     /note="assembly_fragment"
     misc_feature    27226. .28854
                     /note="assembly_fragment"
     misc_feature    28955. .31625
                     /note="assembly_fragment"
     misc_feature    31726. .34084
                     /note="assembly_fragment"
     misc_feature    34185. .36582
                     /note="assembly_fragment"
     misc_feature    36683. .39629
                     /note="assembly_fragment"
     misc_feature    39730. .43900
                     /note="assembly_fragment"
     misc_feature    44001. .46837
                     /note="assembly_fragment"
     misc_feature    46938. .52599
                     /note="assembly_fragment"
     misc_feature    52700. .60235
                     /note="assembly_fragment"
     misc_feature    60336. .67600
                     /note="assembly_fragment"
     misc_feature    67701. .76219
                     /note="assembly_fragment"
     misc_feature    76320. .85003
                     /note="assembly_fragment"
     misc_feature    85104. .95875
                     /note="assembly_fragment"
     misc_feature    95976. .108080
                     /note="assembly_fragment"
     misc_feature    108181. .126996
                     /note="assembly_fragment"
     misc_feature    127097. .145024
                     /note="assembly_fragment"
     misc_feature    145125. .166554
                     /note="assembly_fragment"
     misc_feature    166655. .168177
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    41092 a  41268 c  42342 g  41368 t   2107 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 168177;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 atctgctgtttgaagc 18
           ||||||||||||||||
Db   81387 ATCTGCTGTTTGAAGC 81372


RESULT  32
AP003845
LOCUS       AP003845   171069 bp    DNA             HTG       04-JUL-2001
DEFINITION  Oryza sativa chromosome 7 clone OJ1699_E05, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AP003845
VERSION     AP003845.1  GI:14595190
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1699_E05.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 171069)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
            clone:OJ1699_E05
  JOURNAL   Published Only in Database (2001) In press
REFERENCE   2  (bases 1 to 171069)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Takuji Sasaki, National Institute of
            Agrobiological Resources, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
```

```
     source          1. .171069
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /chromosome="7"
                     /clone="OJ1699_E05"
BASE COUNT    47159 a  38616 c  38364 g  46546 t    384 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 171069;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 atctgctgtttgaagc 18
           ||||||||||||||||
Db 123035 ATCTGCTGTTTGAAGC 123050


RESULT  33
AC090534
LOCUS       AC090534   191677 bp    DNA             HTG       28-JUL-2001
DEFINITION  Mus musculus chromosome 18 clone RP23-233F17 strain C57BL6/J,
            WORKING DRAFT SEQUENCE, 23 unordered pieces.
ACCESSION   AC090534
VERSION     AC090534.7  GI:15027685
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 191677)
  AUTHORS   Li,L., Montgomery,K.T., Grills,G., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Han,J., Ioshikhes,I.P., Lee,E.,
            Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
  TITLE     High Throughput Mouse Sequencing
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191677)
  AUTHORS   Grills,G., Li,L., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Han,J., Ioshikhes,I.P., Lee,E.,
            Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAR-2001) Department of Molecular Genetics, Albert
            Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
            Bronx, NY 10461, USA
COMMENT     On Jul 28, 2001 this sequence version replaced gi:14993657.
            -----------------Genome Center
            Center:  Albert Einstein College of Medicine
            Center Code:  AECOM
            Web site:
            http://sequence.aecom.yu.edu/cgi-
            bin/ws.exe/mouseDB/mouseSEQ/mouseseqtable.hts
            Contact: htgs@sequence.aecom.yu.edu
            -----------------Summary Statistics
            Center project name:  AJT
            Sequencing vector:  pUC18; L08752
            Chemistry:  Dye-terminator Big Dye; 100%
            *Consensus quality:  223066 at least Q20
            *Consensus quality:  213794 at least Q30
            *Consensus quality:  199611 at least Q40
            Estimated insert size:   agarose-FP - N/A
            **Estimated insert size: 191237 - sum-of-contigs
            Quality coverage:  agarose-FP - N/A
            Quality coverage:  6.1 x in Q20 bases; sum-of-contigs estimation
            -----------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    42624: contig of 42624 bp in length
            *    42625    42644: gap of unknown length
            *    42645    93762: contig of 51118 bp in length
            *    93763    93782: gap of unknown length
            *    93783   130149: contig of 36367 bp in length
            *   130150   130169: gap of unknown length
            *   130170   160259: contig of 30090 bp in length
            *   160260   160279: gap of unknown length
            *   160280   160617: contig of 338 bp in length
            *   160618   160637: gap of unknown length
            *   160638   170066: contig of 9429 bp in length
            *   170067   170086: gap of unknown length
            *   170087   177793: contig of 7707 bp in length
            *   177794   177813: gap of unknown length
            *   177814   179001: contig of 1188 bp in length
            *   179002   179021: gap of unknown length
            *   179022   180140: contig of 1119 bp in length
            *   180141   180160: gap of unknown length
            *   180161   180415: contig of 255 bp in length
            *   180416   180435: gap of unknown length
            *   180436   180872: contig of 437 bp in length
            *   180873   180892: gap of unknown length
            *   180893   181101: contig of 209 bp in length
            *   181102   181121: gap of unknown length
            *   181122   182014: contig of 893 bp in length
            *   182015   182034: gap of unknown length
            *   182035   183420: contig of 1386 bp in length
            *   183421   183440: gap of unknown length
            *   183441   185321: contig of 1881 bp in length
            *   185322   185341: gap of unknown length
            *   185342   185549: contig of 208 bp in length
            *   185550   185569: gap of unknown length
            *   185570   186845: contig of 1276 bp in length
            *   186846   186865: gap of unknown length
            *   186866   187015: contig of 150 bp in length
            *   187016   187035: gap of unknown length
            *   187036   187715: contig of 680 bp in length
            *   187716   187735: gap of unknown length
            *   187736   188117: contig of 382 bp in length
            *   188118   188137: gap of unknown length
            *   188138   189263: contig of 1126 bp in length
            *   189264   189283: gap of unknown length
            *   189284   190807: contig of 1524 bp in length
            *   190808   190827: gap of unknown length
            *   190828   191677: contig of 850 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191677
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="18"
                     /clone="RP23-233F17"
                     /sex="male"
     misc_feature    1. .42624
                     /note="assembly_name:Contig108
                     clone_end:SP6
                     vector_side:left"
     misc_feature    42645. .93762
                     /note="assembly_name:Contig107"
     misc_feature    93783. .130149
                     /note="assembly_name:Contig106"
     misc_feature    130170. .160259
                     /note="assembly_name:Contig105
                     clone_end:T7
                     vector_side:right"
     misc_feature    160280. .160617
                     /note="assembly_name:Contig104"
     misc_feature    160638. .170066
                     /note="assembly_name:Contig103"
     misc_feature    170087. .177793
                     /note="assembly_name:Contig102"
     misc_feature    177814. .179001
                     /note="assembly_name:Contig101"
```

```
     misc_feature    179022. .180140
                     /note="assembly_name:Contig100"
     misc_feature    180161. .180415
                     /note="assembly_name:Contig99"
     misc_feature    180436. .180872
                     /note="assembly_name:Contig98"
     misc_feature    180893. .181101
                     /note="assembly_name:Contig97"
     misc_feature    181122. .182014
                     /note="assembly_name:Contig96"
     misc_feature    182035. .183420
                     /note="assembly_name:Contig95"
     misc_feature    183441. .185321
                     /note="assembly_name:Contig94"
     misc_feature    185342. .185549
                     /note="assembly_name:Contig93"
     misc_feature    185570. .186845
                     /note="assembly_name:Contig92"
     misc_feature    186866. .187015
                     /note="assembly_name:Contig91"
     misc_feature    187036. .187715
                     /note="assembly_name:Contig90"
     misc_feature    187736. .188117
                     /note="assembly_name:Contig89"
     misc_feature    188138. .189263
                     /note="assembly_name:Contig88"
     misc_feature    189284. .190807
                     /note="assembly_name:Contig87"
     misc_feature    190828. .191677
                     /note="assembly_name:Contig86"
BASE COUNT    52515 a  43754 c  41414 g  53467 t    527 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 191677;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaa 16
          ||||||||||||||||
Db 156118 GGATCTGCTGTTTGAA 156133


RESULT  34
AC007104/c
LOCUS       AC007104   200000 bp    DNA             HTG       17-JUL-1999
DEFINITION  Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 16
            unordered pieces.
ACCESSION   AC007104
VERSION     AC007104.4  GI:5523795
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200000)
  AUTHORS   Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200000)
  AUTHORS   Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-1999) Department of Genetics, Stanford Human
            Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT     On Jul 17, 1999 this sequence version replaced gi:4757683.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1369: contig of 1369 bp in length
            *     1370     3774: gap of unknown length
            *     3775     4878: contig of 1104 bp in length
            *     4879     7283: gap of unknown length
            *     7284     8555: contig of 1272 bp in length
            *     8556    10960: gap of unknown length
            *    10961    12207: contig of 1247 bp in length
            *    12208    14612: gap of unknown length
            *    14613    16262: contig of 1650 bp in length
            *    16263    18667: gap of unknown length
            *    18668    20176: contig of 1509 bp in length
            *    20177    22581: gap of unknown length
            *    22582    24278: contig of 1697 bp in length
            *    24279    26683: gap of unknown length
            *    26684    28198: contig of 1515 bp in length
            *    28199    30603: gap of unknown length
            *    30604    32359: contig of 1756 bp in length
            *    32360    34764: gap of unknown length
            *    34765    39958: contig of 5194 bp in length
            *    39959    42363: gap of unknown length
            *    42364    57085: contig of 14722 bp in length
            *    57086    59490: gap of unknown length
            *    59491    74905: contig of 15415 bp in length
            *    74906    77310: gap of unknown length
            *    77311    84273: contig of 6963 bp in length
            *    84274    86678: gap of unknown length
            *    86679   118061: contig of 31383 bp in length
            *   118062   120466: gap of unknown length
            *   120467   152965: contig of 32499 bp in length
            *   152966   155370: gap of unknown length
            *   155371   200000: contig of 44630 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
BASE COUNT    38319 a  42757 c  43367 g  37466 t  38091 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 2;  Length 200000;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 atctgctgtttgaagc 18
          ||||||||||||||||
Db 129264 ATCTGCTGTTTGAAGC 129249


RESULT  35
AX120085/c
LOCUS       AX120085   349980 bp    DNA             PAT       11-MAY-2001
DEFINITION  Sequence 1 from Patent EP1108790.
ACCESSION   AX120085 AX114121
VERSION     AX120085.1  GI:14036800
KEYWORDS    .
SOURCE      Corynebacterium glutamicum.
  ORGANISM  Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1  (bases 1 to 349980)
  AUTHORS   Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
            Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
  TITLE     Novel polynucleotides
  JOURNAL   Patent: EP 1108790-A 1 20-JUN-2001;
            KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="Corynebacterium glutamicum"
                     /db_xref="taxon:1718"
                     /note="Seq 1 to long (3.309.400) split in 11, seq 1
                     0.000.001 0.349.980 349.980"
```

```
BASE COUNT    79703 a  91547 c  98381 g  80349 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 6;  Length 349980;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 ctgtttgaagcgcagc 23
          ||||||||||||||||
Db 238128 CTGTTTGAAGCGCAGC 238113


RESULT  36
PPHE7D
LOCUS       PPHE7D        306 bp    DNA             VRL       10-FEB-1999
DEFINITION  Human papillomavirus type 21 DNA for E7 protein, complete cds.
ACCESSION   D50548
VERSION     D50548.1  GI:808883
KEYWORDS    E7 protein.
SOURCE      Human papillomavirus type 21 DNA.
  ORGANISM  Human papillomavirus type 21
            Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
            Papillomavirus.
REFERENCE   1  (bases 1 to 306)
  AUTHORS   Kiyono,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1995) to the DDBJ/EMBL/GenBank databases. Tohru
            Kiyono, Aichi Cancer Center, Research Institute, Laboratory of
            Viral Oncology; 1-1 Kanokoden, Chikusa-ku, Nagoya, Aichi 464, Japan
            (E-mail:h44714u@nucc.cc.nagoya-u.ac.jp, Tel:052-762-6111(ex.8838),
            Fax:052-763-5233)
REFERENCE   2  (bases 1 to 306)
  AUTHORS   Adachi,A., Kiyono,T., Ohashi,M. and Ishibashi,M.
  TITLE     Detection of human papillomavirus type 47 DNA in malignant lesions
            of epidermodysplasia verruciformis with protocols for precise
            typing of the related HPV DNAs
  JOURNAL   Unpublished (1995)
REFERENCE   3  (sites)
  AUTHORS   Adachi,A., Kiyono,T., Hayashi,Y., Ohashi,M. and Ishibashi,M.
  TITLE     Detection of human papillomavirus (HPV) type 47 DNA in malignant
            lesions from epidermodysplasia verruciformis by protocols for
            precise typing of related HPV DNAs
  JOURNAL   J. Clin. Microbiol. 34 (2), 369-375 (1996)
  MEDLINE   96381005
FEATURES             Location/Qualifiers
     source          1. .306
                     /organism="Human papillomavirus type 21"
                     /db_xref="taxon:31548"
     gene            1. .306
                     /gene="E7"
     CDS             1. .306
                     /gene="E7"
                     /note="putative"
                     /codon_start=1
                     /product="E7 protein"
                     /protein_id="BAA09116.1"
                     /db_xref="GI:808884"
                     /translation="MIGKEVTLQDIVLELNELQPEVQPVDLFCEEELPSEQQETEEEL
                     PERTSYKVVTPCGCCKVKLRIFVNATRFAIRTFQNLLFEELQLLCPECRGNCKHGGS"
BASE COUNT       88 a     59 c     81 g     78 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 14;  Length 306;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 atctgctgtttgaag 17
          |||||||||||||||
Db    236 ATCTGCTGTTTGAAG 250


RESULT  37
PEU244276
LOCUS       PEU244276     625 bp    DNA             PLN       11-MAY-1999
DEFINITION  Pringsheimia euphorbiae 5.8S rRNA gene and internal transcribed
            spacers 1 and 2 (ITS1, ITS2), strain CBS 747.71.
ACCESSION   AJ244276
VERSION     AJ244276.1  GI:4808321
KEYWORDS    5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer 1;
            internal transcribed spacer 2; ITS1; ITS2.
SOURCE      Pringsheimia euphorbiae.
  ORGANISM  Pringsheimia euphorbiae
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes et
            Chaetothyriomycetes incertae sedis; Dothioraceae; Pringsheimia.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Yurlova,N., Hoog,G.D. and Gerrits van den Ende,A.
  TITLE     Taxonomy of Aureobasidium and allied genera
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 625)
  AUTHORS   Hoog,G.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-1999) Hoog G.D., Research, Centraalbureau voor
            Schimmelcultures, P.O.Box 273, 3740 AG Baarn, NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Pringsheimia euphorbiae"
                     /strain="CBS 747.71"
                     /db_xref="taxon:93493"
     misc_feature    <1. .196
                     /note="internal transcribed spacer 1, ITS1"
     rRNA            197. .351
                     /gene="5.8S rRNA"
                     /product="5.8S ribosomal RNA"
     gene            197. .351
                     /gene="5.8S rRNA"
     misc_feature    352. .>625
                     /note="internal transcribed spacer 2, ITS2"
BASE COUNT      170 a    128 c    135 g    192 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 8;  Length 625;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     10 gtttgaagcgcagca 24
          |||||||||||||||
Db    438 GTTTGAAGCGCAGCA 452


RESULT  38
AX110623
LOCUS       AX110623      810 bp    DNA             PAT       30-APR-2001
DEFINITION  Sequence 1356 from Patent WO0123604.
ACCESSION   AX110623
VERSION     AX110623.1  GI:13926915
KEYWORDS    .
SOURCE      Serratia marcescens.
  ORGANISM  Serratia marcescens
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Serratia.
REFERENCE   1  (bases 1 to 810)
  AUTHORS   Bergeron,M.G., Boissinot,M., Huletsky,A., m Nard,C., Ouellette,M.,
            Picard,F.J. and Roy,P.H.
  TITLE     Highly conserved genes and their use to generate probes and primers
            for detection of microorganisms
  JOURNAL   Patent: WO 0123604-A 1356 05-APR-2001;
            Infectio Diagnostic (I.D.I.) INC. (CA)
FEATURES             Location/Qualifiers
     source          1. .810
                     /organism="Serratia marcescens"
                     /db_xref="taxon:615"
BASE COUNT      129 a    270 c    268 g    143 t
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 6;  Length 810;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 ctgtttgaagcgcag 22
          |||||||||||||||
Db    742 CTGTTTGAAGCGCAG 756


RESULT  39
AF128956
LOCUS       AF128956     1551 bp    DNA             BCT       21-JUN-1999
DEFINITION  Escherichia coli strain E2348/69 flagellin (fliC) gene, partial
            cds.
ACCESSION   AF128956
VERSION     AF128956.2  GI:5107854
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1551)
  AUTHORS   Reid,S.D., Selander,R.K. and Whittam,T.S.
  TITLE     Sequence diversity of flagellin (fliC) alleles in pathogenic
            Escherichia coli
  JOURNAL   J. Bacteriol. 181 (1), 153-160 (1999)
  MEDLINE   99084952
REFERENCE   2  (bases 1 to 1551)
  AUTHORS   Reid,S.D. and Whittam,T.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-1999) Biology Department, Institute of Molecular
            Evolutionary Genetics, The Pennsylvania State University, 208
            Mueller Laboratory, University Park, PA 16802, USA
REFERENCE   3  (bases 1 to 1551)
  AUTHORS   Reid,S.D. and Whittam,T.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-1999) Biology Department, Institute of Molecular
            Evolutionary Genetics, The Pennsylvania State University, 208
            Mueller Laboratory, University Park, PA 16802, USA
  REMARK    Sequence update by submitter
COMMENT     On Jun 21, 1999 this sequence version replaced gi:4731615.
FEATURES             Location/Qualifiers
     source          1. .1551
                     /organism="Escherichia coli"
                     /strain="E2348/69"
                     /db_xref="taxon:562"
                     /note="enteropathogenic"
     gene            <1. .>1551
                     /gene="fliC"
     CDS             <1. .>1551
                     /gene="fliC"
                     /codon_start=1
                     /transl_table=11
                     /product="flagellin"
                     /protein_id="AAD28527.2"
                     /db_xref="GI:5107855"
                     /translation="QSALSSSIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQAA
                     RNANDGISVAQTTEGALSEINNNLQRIRELTVQASTGTNSDSDLDSIQDEIKSRLDEI
                     DRVSGQTQFNGVNVLAKDGSMKIQVGANDGQTITIDLKKIDSDTLGLNGFNVNGKGET
                     ANTAATLKDMSGFTAAAAPGGTVGVTQYTDKSAVASSVNILNAVAGANGNKVTTSADV
                     GFGTPAAAVTYTYNKDTNSYSAASDDISSANLAAFLNPQARDTTKATVTIGGKDQDVN
                     IDKSGNLTAADDGAVLYMDATGNLTKNNAGGDTQATLAKVATATGAKAATIQTDKGTF
                     TSDGTAFDGASMSIDANTFANAVKNDTYTATVGAKTYSVTTGSAAADTAYMSNGVLSD
                     TPPTYYAQADGSITTTEDAAAGKLVYKGSDGKLTTDTTSKAESTSDPLAALDDAISQI
                     DKFRSSLGAVQNRLDSAVTNLNNTTTNLFEAQSRIQDADYATEVSNMSKAQIIQQAGN
                     SVLAKANQV"
BASE COUNT      443 a    375 c    368 g    365 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 1;  Length 1551;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 ctgtttgaagcgcag 22
          |||||||||||||||
Db   1432 CTGTTTGAAGCGCAG 1446


RESULT  40
SMAAAC3VB
LOCUS       SMAAAC3VB    1572 bp    DNA             BCT       26-APR-1993
DEFINITION  Serratia marcescens N-acetyltransferase (aac3-Vb) gene, complete
            cds.
ACCESSION   M97172
VERSION     M97172.1  GI:152811
KEYWORDS    acetyltransferase.
SOURCE      Serratia marcescens DNA.
  ORGANISM  Serratia marcescens
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Serratia.
REFERENCE   1  (bases 1 to 1572)
  AUTHORS   Rather,P.N., Mierzwa,R., Hare,R., Miller,G. and Shaw,K.
  TITLE     Cloning and DNA sequence analysis of an aac(3)-Vb gene from
            Serratia marcescens
  JOURNAL   Antimicrob. Agents Chemother. 36, 2222-2227 (1992)
  MEDLINE   93073766
FEATURES             Location/Qualifiers
     source          1. .1572
                     /organism="Serratia marcescens"
                     /db_xref="taxon:615"
     gene            656. .1465
                     /gene="aac3-Vb"
     CDS             656. .1465
                     /gene="aac3-Vb"
                     /codon_start=1
                     /transl_table=11
                     /product="acetyltransferase"
                     /protein_id="AAA26548.1"
                     /db_xref="GI:152812"
                     /translation="MNTIESITADLHGLGVRPGDLIMVHASLKAVGPVEGGAASVVSA
                     LRAAVGSAGTLMGYASWDRSPYEETLNGARMDEELRRRWPPFDLATSGTYPGFGLLNR
                     FLLEAPDARRSAHPDASMVAVGPLAATLTEPHRLGQALGEGSPLERFVGHGGKVLLLG
                     APLDSVTVLHYAEAIAPIPNKRRVTYEMPMLGPDGRVRWELAEDFDSNGILDCFAVDG
                     KPDAVETIAKAYVELGRHREGIVGRAPSYLFEAQDIVSFGVTYLEQHFGAP"
BASE COUNT      276 a    516 c    506 g    274 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 1;  Length 1572;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 ctgtttgaagcgcag 22
          |||||||||||||||
Db   1397 CTGTTTGAAGCGCAG 1411


RESULT  41
AF239157/c
LOCUS       AF239157     1616 bp    mRNA            ROD       06-JUL-2001
DEFINITION  Rattus norvegicus DEXRAS1 (Dexras1) mRNA, complete cds.
ACCESSION   AF239157
VERSION     AF239157.1  GI:7230767
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1616)
  AUTHORS   Fang,M., Jaffrey,S.R., Sawa,A., Ye,K., Luo,X. and Snyder,S.H.
  TITLE     Dexras1: a G protein specifically coupled to neuronal nitric oxide
            synthase via CAPON
```

```
  JOURNAL   Neuron 28 (1), 183-193 (2000)
  MEDLINE   20537828
   PUBMED   11086993
REFERENCE   2  (bases 1 to 1616)
  AUTHORS   Fang,M., Jaffrey,S.R., Sawa,A., Ye,K. and Snyder,S.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-2000) Neuroscience, Johns Hopkins University, 725
            N. Wolfe St, Baltimore, MD 21205, USA
FEATURES             Location/Qualifiers
     source          1. .1616
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            1. .1616
                     /gene="Dexras1"
     CDS             118. .960
                     /gene="Dexras1"
                     /note="Ras-related protein"
                     /codon_start=1
                     /product="DEXRAS1"
                     /protein_id="AAF43090.1"
                     /db_xref="GI:7230768"
                     /translation="MKLAAMIKKMCPSDSELSIPAKNCYRMVILGSSKVGKTA1VSRF
                     LTGRFEDAYTPTIEDFHRKFYSIRGEVYQLDILDTSGNHPFPAMRRLSILTGDVFILV
                     FSLDNRDSFEEVQRLKQQILDTKSCLKNKTKENVDVPLVICGNKGDRDFYREVEQREI
                     EQLVGDDPQRCAYFEISAKKNSSLDQMFRALFAMAKLPSEMSPDLHRKVSVQYCDVLH
                     KKALRNKKLLRAGSGGGGDHGDAFGILAPFARRPSVHSDLMYIREKTSVSSQAKDKER
                     CVIS"
BASE COUNT      403 a    436 c    413 g    363 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 1616;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttga 15
          ||||||||||||||||
Db    481 GGATCTGCTGTTTGA 467


RESULT  42
AF009246/c
LOCUS       AF009246     1623 bp    mRNA            ROD       13-FEB-1998
DEFINITION  Mus musculus ras-related protein (DEXRAS1) mRNA, complete cds.
ACCESSION   AF009246
VERSION     AF009246.1  GI:2253712
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1623)
  AUTHORS   Kemppainen,R.J. and Behrend,E.N.
  TITLE     Dexamethasone rapidly induces a novel ras superfamily
            member-related gene in AtT-20 cells
  JOURNAL   J. Biol. Chem. 273 (6), 3129-3131 (1998)
  MEDLINE   98123070
REFERENCE   2  (bases 1 to 1623)
  AUTHORS   Kemppainen,R.J. and Behrend,E.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUN-1997) Physiology & Pharmacology, Auburn
            University College of Veterinary Medicine, 213 Greene Hall, Auburn,
            AL 36849, USA
FEATURES             Location/Qualifiers
     source          1. .1623
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_line="corticotrope tumor cell line AtT-20"
     gene            1. .1623
                     /gene="DEXRAS1"
     CDS             142. .984
                     /gene="DEXRAS1"
                     /note="induced by dexamethasone"
                     /codon_start=1
                     /product="ras-related protein"
                     /protein_id="AAC53538.1"
                     /db_xref="GI:2253713"
                     /translation="MKLAAMIKKMCPSDSELSIPAKNCYRMVILGSSKVGKTAIVSRF
                     LTGRFEDAYTPTIEDFHRKFYSIRGEVYQLDILDTSGNHPFPAMRRLSILTGDVFILV
                     FSLDNRDSFEEVQRLKQQILDTKSCLKNKTKENVDVPLVICGNKGDRDFYREVEQREI
                     EQLVGDDPQRCAYFEISAKKNSSLDQMFRALFAMAKLPSEMSPDLHRKVSVQYCDVLH
                     KKALRNKKLLRAGSGGGGDHGDAFGILAPFARRPSVHSDLMYIREKTSVGSQAKDKER
                     CVIS"
BASE COUNT      412 a    437 c    417 g    357 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 1623;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttga 15
          ||||||||||||||||
Db    505 GGATCTGCTGTTTGA 491


RESULT  43
AF039572
LOCUS       AF039572     2795 bp    DNA             BCT       02-DEC-1998
DEFINITION  Serratia liquefaciens serrawettin synthase (swrA) gene, partial
            cds.
ACCESSION   AF039572
VERSION     AF039572.1  GI:2760927
KEYWORDS    .
SOURCE      Serratia liquefaciens.
  ORGANISM  Serratia liquefaciens
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Serratia.
REFERENCE   1  (bases 1 to 2795)
  AUTHORS   Lindum,P.W., Anthoni,U., Christophersen,C., Eberl,L., Molin,S. and
            Givskov,M.
  TITLE     N-Acyl-L-homoserine lactone autoinducers control production of an
            extracellular lipopeptide biosurfactant required for swarming
            motility of Serratia liquefaciens MG1
  JOURNAL   J. Bacteriol. 180 (23), 6384-6388 (1998)
  MEDLINE   99047587
REFERENCE   2  (bases 1 to 2795)
  AUTHORS   Lindum,P.W., Eberl,L., Anthoni,U., Christophersen,C., Molin,S. and
            Givskov,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-1997) Dept. of Microbiology, Tech. University of
            Denmark, Bldg. 301, Lyngby DK-2800, Denmark
FEATURES             Location/Qualifiers
     source          1. .2795
                     /organism="Serratia liquefaciens"
                     /strain="MG1"
                     /db_xref="taxon:614"
     gene            111. .>2795
                     /gene="swrA"
     CDS             111. .>2795
                     /gene="swrA"
                     /function="produces the lipodepsipeptide serrawettin W2"
                     /note="serrawettin W2 is required for the rapid surface
                     translocation known as swarming motility."
                     /codon_start=1
                     /transl_table=11
                     /product="serrawettin synthase"
                     /protein_id="AAC82333.1"
                     /db_xref="GI:2760928"
                     /translation="MLRWRKSNCMKLPPLQRELWLSASMSEGSEGLPFNMWGYGTIEG
                     PLQIALLRQAIDAIIEETPAMQAHFTEQDGELYQYRAARSHDPLSAHDFSSQEYPELA
                     AQRWMRQDADRVISALDQDLFGFALLTLAPGSYILYRRFHHMLVDGRSAEEFMRRIAL
                     YYNALAEKKSIPEFSNCSFSRLYEHDLQYRDSSRFATDKQFWQGYTLAAPHSLTRQEK
                     LVPYSSMQISNHTLTINEIAAIDRGAEKAGVHRAHILAAAVALCFHSITGQTLLNFSL
                     PVTGTRERNSIGMTANVVPLILAINPESRLADFAQQVAAEIAKVVRRQLYRGEDIRRD
                     RSTMSSSWFGPAINIVSFDHGDPFWGCKTRWYYGGNIAVSDMQILFYEDQQAQELDVM
```

```
                     FADAPHAHSLSQLESLQRRFQFILRTLSEQPEGSIAQVMQQIHAQPEADVSGSFYDLR
                     RLPPAAGFIRWDRPAAELVRLATSLGHPAGDCPLKILLADRALGVGRLEMLPDAGDAA
                     PGTLLGVDDAGWNIATAEGAVRIGGWFTAEGELAAGEALARRHGLTVGEPLPLLDPLQ
                     VKRLEEIQAAGGHHAAFWQPRLAQFAPAAFPVAWQRETRPAWQATAWQHAGKIEVESV
                     LTGLALYLGLMGGESDFQLGWLNATDDVIGESGQVWTLPLHISLNPTLGFTEAQVQMR
                     REWQQAYARRAYHPALLGLDTRSRAWSTVVALLRRDTSPGFDEGQEGAAAVKRSGARA
                     VLQICESNGAFRWVYDSAAVGDEQMLRAAQHLLLLLRQAAAAPNAPVRGDRLLSQAER
                     ELLLNVRNRTHRAFPQASSVSALFEAQARRTPTAPALSQGETSLSYAELNARANRLAH
                     HLTARGVGPESRVAVCAARSPEMIVALLAILKAGGAYVP"
BASE COUNT      545 a    786 c    888 g    576 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 1;  Length 2795;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 ctgtttgaagcgcag 22
           |||||||||||||||
Db    2571 CTGTTTGAAGCGCAG 2585

.

RESULT  44
PMO238607/c
LOCUS       PMO238607    6043 bp    RNA             VRL       15-OCT-1999
DEFINITION  Potato mop-top virus RNA1 complete genome, Swedish isolate (Sw).
ACCESSION   AJ238607
VERSION     AJ238607.1  GI:6066602
KEYWORDS    polymerase.
SOURCE      Potato mop-top virus.
  ORGANISM  Potato mop-top virus
            Viruses; ssRNA positive-strand viruses, no DNA stage; Furovirus.
REFERENCE   1  (bases 1 to 6043)
  AUTHORS   Savenkov,E.I., Sandgren,M. and Valkonen,J.P.
  TITLE     Complete sequence of RNA 1 and the presence of tRNA-like structures
            in all RNAs of Potato mop-top virus, genus Pomovirus
  JOURNAL   J. Gen. Virol. 80 (Pt 10), 2779-2784 (1999)
  MEDLINE   20037846
REFERENCE   2  (bases 1 to 6043)
  AUTHORS   Savenkov,E.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1999) Savenkov E.I., Plant Biology, Slu, box
            7080, se-750 07, Uppsala, SWEDEN
FEATURES             Location/Qualifiers
     source          1. .6043
                     /organism="Potato mop-top virus"
                     /virion
                     /isolate="Swedish (Sw)"
                     /db_xref="taxon:37128"
                     /country="Sweden"
                     /note="RNA1 RNA; genus Pomovirus"
     CDS             115. .4026
                     /function="replication"
                     /note="ORF1; readthrough protein"
                     /codon_start=1
                     /product="polymerase"
                     /protein_id="CAB58364.1"
                     /db_xref="GI:6066603"
                     /translation="MDQSILSKITSDDYVNAILHTSATRVGSDLHNAMCNVIVDQIKD
                     STEKNKKKKKIDVKRNLNEDQIQLLAELFPERRVVTSSVHRGTHSMAAAMRKIETDVI
                     FTSFPKNGVIYDIGGNWATHAKRDDGGFVHCCCPILDFRDAQRKMTRLIDFNRFIDDA
                     KVVSAEKAAVAAQIKKDCDLISENAKKDAYDANDLNGTWFCQNKFEDCVYDHSTLGDG
                     KKEAYGMAIHSIYDIHLVDLVSAMERKKVRVLKGTFLFSADIIIGKKRGELPSVNGFY
                     IIDGESIKYSFYDDPNCGYEHNLNSLMLYVTKTFVKAAGGAVYYLELTEMRGDTMFFT
                     ITDASEARVMGVLHDSSTKCLPLNKRDLVVFPLFDIDRATDELVFREELLSREFVNRA
                     LEYAFQLKDNQVTAEGLISYFASTNNAVVIGGSARKTSEKVDPKLLPMITTTLMVYQE
                     LQKAKQKRVLGKLKSKVKEELTLSGILLESVVHRVFGRQSLYQRGLGVFAKWMQYSYGQ
                     DLVGIHDVPLYLEINDRIKLGSALKNVNGFSLSFSELDEKVSLYEEYERERQRISDEI
                     VSEKIGLIGEGYVKVGESSLKPKQKVASRDGMAQWVSGECHLYNTMKCEEKPVEKPRR
                     FSKVVLEEWISEGNCFALNNSFGDEVHWFDSLKEACGRAWRKQVSAVTFSDAEYFDNI
                     ENTENEELEEEEVIKTTEQVEQDWPIGNLPDVDPDDSASAQVCLTESASTSSDEDCDP
                     MEQLIVAACERAFATKKFEEVTMDSQVVQSEEIEIVREGEHAHANLSSGESESSSQSQ
                     ELVAVGSVPLSKWAQMVEDSEVRARQCAVDHDLSWDEVKYAKMPERPEEAEGDDFRTK
                     AKREFLWYLKCKLVADKSTLVEIMRDFIYGQFHSGACETPKNACFLSYEKNVCGEWMF
                     GKRYRHPSKGASSYAVRFTREDWKRAKLIKLQWKNAKAEENGEMSSDNSDKPIVPQGE
                     TGIYLFCDITFLMNEIPILNRLEISFKKRVQRRAPRITLVDGVPGCGKSTYVVKEANL
                     VNQYVVTIGREAAEDLRERFKSERNATATQLKRVRTVDSYLLNDTQSRANVLHFDEAL
                     MAHAGMVYFCADDLSARSVICQGDSQQIPFINRVESITLRYSKLEIDNVVEKRLTYRS
                     PLDVASYLTKKNFYGTSVVTSANPLVRSLKTVGPRDGMTSIYSIPKIPGTQYLTFLQS
                     EKEEMRQYLGRGNWNVNTVHESQGKTYDNVVLCRLKATDNEIYPGGRNSSPYMVVGVT
                     RHRRSLVYYTKAEDKLYFDLAEMLSVQEGKLMKHLHEEGVK"
     mat_peptide     118. .4023
                     /product="polymerase"
BASE COUNT     1791 a    935 c   1578 g   1739 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 14;  Length 6043;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 ctgtttgaagcgcag 22
           |||||||||||||||
Db    4645 CTGTTTGAAGCGCAG 4631


RESULT  45
HPU31779
LOCUS       HPU31779     7779 bp    DNA             VRL       17-OCT-1995
DEFINITION  Human papillomavirus type 21, complete genome.
ACCESSION   U31779
VERSION     U31779.1  GI:1020170
KEYWORDS    .
SOURCE      Human papillomavirus type 21.
  ORGANISM  Human papillomavirus type 21
            Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
            Papillomavirus.
REFERENCE   1  (bases 1 to 7779)
  AUTHORS   Delius,H.
  TITLE     Sequenced by Hajo Delius, Deutsches Krebsforschungzentrum,
            Angewandte Tumorvirologie, I.N.F. 506, W-6900 Heidelberg, Germany
  JOURNAL   Unpublished
REFERENCE   2  (sites)
  AUTHORS   Kremsdorf,D., Favre,M., Jablonska,S., Obalek,S., Rueda,L.A.,
            Lutzner,M.A., Blanchet-Bardon,C., Van Voorst Vader,P.C. and Orth,G.
  TITLE     Molecular cloning and characterization of the genomes of nine newly
            recognized human papillomavirus types associated with
            epidermodysplasia verruciformis
  JOURNAL   J. Virol. 52 (3), 1013-1018 (1984)
  MEDLINE   85033930
REFERENCE   3  (sites)
  AUTHORS   Kiyono,T., Hiraiwa,A. and Ishibashi,M.
  TITLE     Differences in transforming activity and coded amino acid sequence
            among E6 genes of several papillomaviruses associated with
            epidermodysplasia verruciformis
  JOURNAL   Virology 186 (2), 628-639 (1992)
  MEDLINE   92124737
REFERENCE   4  (bases 1 to 7779)
  AUTHORS   Farmer,A.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-1995) Andrew D. Farmer, HIV Sequence Database,
            Los Alamos National Labroatory, T-10, Mail Stop K710, Los Alamos,
            NM 87501, USA
COMMENT     HPV21 was originally isolated from skin warts of an
            epidermodysplasia verruciformis (EV) patient [2]. Hybridization
            assays and phylogenetic reconstructions based on DNA sequences
            indicate that HPV21 is most closely related to HPV14 and HPV20, and
            then to HPV19 and HPV25.  This grouping agrees with assays of the
            degree of transforming activity of the E6 gene (these related HPV
            types had relatively low transforming activity as compared to HPVs
            5, 8, and 47), and clustering of similarity of amino acids in the
            second zinc finger domain of E6 [3]. The E6 gene of HPVs 14, 21,
            and 25 can enhance the induction of anchorage independent growth of
            3Y1 cells by the HPV16 E7 gene, although again less effectively
            than that of HPVs 5, 8, and 47.
FEATURES             Location/Qualifiers
```

```
source          1. .7779
                /organism="Human papillomavirus type 21"
                /db_xref="taxon:31548"
                /note="cloned HPV-21 was obtained from the Papillomavirus
                Reference Center, Heidelberg and sequenced by Dr. H.
                Delius"
gene            200. .706
                /gene="E6"
CDS             200. .706
                /gene="E6"
                /note="putative"
                /codon_start=1
                /product="transforming protein E6"
                /protein_id="AAA79394.1"
                /db_xref="GI:1020171"
                /translation="MADSSTDSADEGPSPKRRHLEEENTSSFLEPPLPATIRDLANLL
                EIPLDDCLVPCNFCGNFLTHLEVCEFDEKKLSLLWKDHCVFACCRVCCAATATYEYNE
                FYESTVVGRDIEEITGKSIFDIDVRCYNCMKFLDSIEKLDICGRKFFFHKVRGSWKGI
                CRLCKHFQ"
gene            706. .1011
                /gene="E7"
CDS             706. .1011
                /gene="E7"
                /note="putative"
                /codon_start=1
                /product="transforming protein E7"
                /protein_id="AAA79395.1"
                /db_xref="GI:1020172"
                /translation="MIGKEVTLQDIVLELNELQPEVQPVDLFCEEELPSEQQETEEEL
                PERTAYKVVTPCGCCKVKLRIFVNATQFAIRTFQNLLFEELQLLCPECRGNCKHGGS"
gene            998. .2809
                /gene="E1"
CDS             998. .2809
                /gene="E1"
                /note="putative"
                /codon_start=1
                /product="replication protein E1"
                /protein_id="AAA79396.1"
                /db_xref="GI:1020173"
                /translation="MADPKGSTSKEGLEDWCIVEAECSDVENDLEELFDRDTDSDISE
                LLDDNDLEQGNSRELFHQQESKESEEQLQKLKRKYLSPKAVAQLSPRLESITLSPQQK
                SKRRLFAEQDSGLECTLTNEEDVSSEVEVPALDSQPVAEAQLGTVDIHYKELLRASNN
                KAILMAKFKEFFGVGFNDLTRQFKSYKTCCNAWVLSVYAVHDDLLESSKQLLQQHCDY
                IWIRGIGAMSLFLLCFKVGKNRGTVHKLMTAMLNVHEKQIISEPPKLRNVAAALFWYK
                GAMGSGAFTYGPYPDWIAQQTIVGHQSTEASAFDMSAMVQWAFDNNYLDEADIAYQYA
                KLAPEDSNAVAWLAHNNQARYVREVASMVRFYKKGQMKEMSMSEWIHTRINEVEGEGH
                WSTIAKFLRYQQVNFIMFLAALKDMLHSVPKRNCILIYGPPNTGKSAFTMSLIHVLRG
                RVLSFVNSKSQFWLQPMSECKIALIDDVTDPCWIYMDTYLRNGLDGHVVSLDCKHKAP
                MQTKFPALLLTSNINVHNEVNYRYLHSRIKGFEFPNPFPMKADNTPEFELTDQSWKSF
                FTRLWNQLELSDQEDEGENGESQRSFQCSARSANEHL"
gene            2751. .4262
                /gene="E2"
CDS             2751. .4262
                /gene="E2"
                /note="putative"
                /codon_start=1
                /product="regulatory protein E2"
                /protein_id="AAA79397.1"
                /db_xref="GI:1020174"
                /translation="MENLSDRFNVLQDQLMNIYESAANTIESQIEHWQTLRKEAVLLY
                FARQKGVTRLGYQYVPPLAVSESRAKQAIGMMLQLQSLQKSEYAKEPWSLVDTSAETF
                RSPPENHFKKGPVSVEVIYDNDKDNANAYTMWRYVYYVDDDDQWHKSPSGVNHTGIYF
                MQGTFRHYYVLFADDASRYSRTGHWEVNVNKETVFAPVTSSTPPDSPGGQADSNTSST
                TPATTTDSTSRLSSTRKQSQQTNTKGRRYGRRPSSRTRRTTQTHQRRRSRSKSRSRSR
                SRSRLRSRSRSRSRSYSRSRSQSSDQPQYRFRSGGQVSLITTATTTTTTATNYSTRGS
                GRGSSSTSSSTSKRPRRPRGGAIGGSSGRGRRSSSTSPSPSKRSRGKSESVRQRGISP
                DDVGKSLQSVSTRNTGRLGRLLDEALDPPVILVRGEPNTLKCFRNRAKLKYAGLYKAF
                STAWSWVAGDGTERLGRSRMLISFFSFEQRKDFDKTVKYPKGVDRSYGSFDSL"
gene            3037. .4017
                /gene="E4"
CDS             <3037. .4017
                /gene="E4"
                /note="putative"
                /codon_start=1
                /product="transforming protein E4"
                /protein_id="AAA79398.1"
                /db_xref="GI:1020175"
                /translation="IPVQRHLEALLKIISKKGQCQLRLFMITIKTMLMLTPCGDMFIT
                WMMTTNGIKVQAVSTTQAYILCKELLDTTMFYLLMMQVDIAELDIGKLTLIRKLCLLL
                SPAPPHPTHQEDKQTQTPPPRPPPPPLTPRPDSRPPENSHNKPTPKEEGTDGDRPVGP
                GERPKRIKGGDRGPSPGRGRGRGRGSDPDPGPDPGPIPGPGLNRLTSRNTDSDPEGKC
                PSSLPPPPPPPPQPTTPPEGQGEGHPPPPPPPPNGHDGHEEGPLEGAVGGGDGHPPPP
                PAPPNGHEESQSLLGNVASLLTTWESLFNQLVQEIQVDLEDYWTKLSIPQ"
gene            4351. .5913
                /gene="L2"
CDS             4351. .5913
                /gene="L2"
                /note="putative"
                /codon_start=1
                /product="minor capsid protein L2"
                /protein_id="AAA79399.1"
                /db_xref="GI:1020176"
                /translation="MARAKRVKRDSATNIYRTCKQAGTCPPDVINKVESTTIADKILQ
                YGSAGVFFGGLGISTGKGTGGTTGYVPLGEGPAVRVGNAPTVIRPALVPDTIGPSDII
                PVDTLNPVEPTTSSIVPLTDSTGPDLLPGEVETIAEIHPGPTRPPPDTAVTTSTNGSS
                AVLEVAPEPTPPSRVRVTRTQYHNPSFQVITESTPTTGESSLADHILVTSGTGGQTIG
                GSTPELIELQDFPSRYSFEIEEPTPPRRTSTPIQRIQNIIRRRGGGLTNRRLVQQVNV
                ENPLFVSRPSRLVQFQFDNPAFEEEVTQIFEQDIDTFNEPPDRDFLDIKTLGRPQYSE
                TPAGYVRVSRLGKRGTIRTRSGTQIGSQVHFYRDLSTINTEDPIELQLLGEHSGDATI
                VQGPVESTFIDINVDENPLSEDFSAHSDDLLLDEANEDFSGSQLVVGGRRSTSSYTVP
                RFETTRSGSYYVQDTKGYYVAYPEDRDTSTDIIYPTPDLPVVIIHTFDTSGDFYLHPS
                LSRKFKRRRKYL"
gene            5929. .7485
                /gene="L1"
CDS             5929. .7485
                /gene="L1"
                /note="putative"
                /codon_start=1
                /product="major capsid protein L1"
                /protein_id="AAA79400.1"
                /db_xref="GI:1020177"
                /translation="MAVWQAASGKVYLPPSTPVARVQSTDEYVQRTNIYYHAYSDRLL
                TVGHPYFNVYDVNSAKIKVPKVSGNQHRVFRLKLPDPNRFALADMSVYNPDKERLVWA
                CRGIEIGRGQPLGVGSVGHPLFNKVGDTENPSSYKTQPNSTDDRQNVSFDPKQLQMFI
                IGCAPCLGEHWDKAIPCATDNPPPGSCPPIELINSAIQDGDMADIGYGNLNFKALQQN
                RSDVSLDIVNETCKYPDFLKMQNDVYGDSCFFYARREQCYARHFFVRGGKTGDDIPAG
                QIDEGSMKNAYYIPPMNDQAQYKIGNSMYFPTVSGSLVSSDAQLFNRPFWLQRAQGHN
                NGICWFNQLFVTVVDNTRNTNFSISVNPENADVSKIENYKAESFQEYLRHVEEYELSL
                ILQLCKVPLTAEVLAQINAMNANILEEWQLGFVPAPDNPIHDTYRYIDSAATRCPDKN
                PPKEREDPYKNMKFWDVDLTERLSLDLDQYSLGRKFLFQAGLQQTTVNGTKTLSSRVS
                TRGIKRKRKN"
BASE COUNT      2426 a   1518 c   1680 g   2155 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 14;  Length 7779;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 atctgctgtttgaag 17
           |||||||||||||||
Db     941 ATCTGCTGTTTGAAG 955



Search completed: December 26, 2001, 12:51:20
Job time: 8802 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:16:34 ; Search time 368.09 Seconds
(without alignments)
58.228 Million cell updates/sec

Title: US-09-396-196F-4
Perfect score: 25
Sequence: 1 ggatctgctgtttgaagcgcagcag 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 9

Total number of hits satisfying chosen parameters: 50995

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25	100.0	839	22	AAI62941	Human genomic DNA
2		25	100.0	1041	20	AAX01303	E. coli biotin syn
3		25	100.0	1084	10	AAN91329	E.coli Bio B gene.
4		25	100.0	1121	7	AAN60496	Sequence encoding
5		25	100.0	5872	15	AAQ62386	Biotin-biosynthesi
6	c	17	68.0	5504	20	AAX20559	Polynucleotide seq
7	c	16	64.0	801	22	AAH65216	C glutamicum codin
8	c	16	64.0	924	22	AAF71927	Corynebacterium gl
9	c	16	64.0	349980	22	AAH64966	C glutamicum codin
10		15	60.0	810	22	AAH01365	Serratia marcescen
11	c	15	60.0	1689	20	AAZ23024	Rat kd312 polypept
12	c	15	60.0	2022	21	AAC44132	Zea mays DNA fragm
13	c	15	60.0	3079	20	AAZ23025	Rat kd312 genomic
14	c	14	56.0	336	22	AAF65487	Novel human polynu
15	c	14	56.0	379	20	AAV87319	EST clone BP797.
16		14	56.0	393	22	AAH10789	Human cDNA clone (
17		14	56.0	469	22	AAI12900	Probe #2833 for ge
18		14	56.0	469	22	AAI34258	Probe #2944 used t
19		14	56.0	469	22	AAI02819	Probe #2810 used t
20		14	56.0	476	22	AAI22139	Probe #12072 for g
21		14	56.0	476	22	AAI47435	Probe #16121 used
22		14	56.0	476	22	AAI07839	Probe #7830 used t
23	c	14	56.0	619	21	AAA44124	Human secreted exp
24	c	14	56.0	753	20	AAZ00347	Nucleotide sequenc
25	c	14	56.0	753	22	AAF89005	Human FATP3 coding
26	c	14	56.0	1173	20	AAZ00358	Nucleotide sequenc
27	c	14	56.0	1173	22	AAF89020	Human FATP3 partia
28	c	14	56.0	1244	21	AAA50896	Soybean PI metabol
29		14	56.0	1536	11	AAQ04684	SalI/KpnI fragment
30	c	14	56.0	1699	22	AAI58178	Human polynucleoti
31	c	14	56.0	1998	20	AAZ00367	Nucleotide sequenc
32	c	14	56.0	1998	22	AAF89029	Murine FATP3 parti
33	c	14	56.0	2087	20	AAZ00343	Nucleotide sequenc
34	c	14	56.0	2087	22	AAF89001	Murine FATP3 codin
35	c	14	56.0	2166	20	AAZ00379	Nucleotide sequenc
36	c	14	56.0	2166	22	AAF89041	Human FATP3 coding
37	c	14	56.0	2314	22	AAH17621	Human cDNA sequenc
38	c	14	56.0	2319	22	AAH19223	Human secreted pro
39	c	14	56.0	2339	22	AAD07655	Human secreted pro
40	c	14	56.0	2402	22	AAF89054	Human FATP3 coding
41	c	14	56.0	2405	22	AAF27417	Human fatty acid t
42	c	14	56.0	2447	22	AAH19186	Human secreted pro
43	c	14	56.0	2574	20	AAZ33977	Human PRO703 nucle
44	c	14	56.0	2574	21	AAC78481	Human PRO703 (UNQ3
45	c	14	56.0	2574	21	AAC58239	Human PRO703 nucle
46		14	56.0	2628	22	AAH66254	C glutamicum codin
47	c	14	56.0	2669	20	AAZ25020	Human H-beta 58 fa
48	c	14	56.0	2675	21	AAC98021	Human colon cancer
49		14	56.0	2845	22	AAH18012	Human cDNA sequenc
50	c	14	56.0	7778	18	AAV74359	Staphylococcus aur
51		14	56.0	12284	11	AAQ06001	Sequence encoding
52		14	56.0	12284	15	AAQ71308	Sequence encoding
53		14	56.0	12284	17	AAT29591	Hog cholera virus.
54		14	56.0	12284	22	AAF30678	Hog cholera virus
55		14	56.0	63563	22	AAF28546	Genomic fragment #
56	c	14	56.0	124884	22	AAH74201	Nucleotide sequenc
57	c	14	56.0	125157	22	AAH74202	Nucleotide sequenc
58		14	56.0	349980	22	AAH68527	C glutamicum codin
59		14	56.0	349980	22	AAH68528	C glutamicum codin
60	c	13	52.0	17	16	AAQ94360	Septoria nodorum I
61	c	13	52.0	17	19	AAV62508	Septoria nodorum s
62	c	13	52.0	23	16	AAQ94361	Septoria nodorum I
63	c	13	52.0	23	19	AAV62509	Septoria nodorum s
64	c	13	52.0	80	22	AAI50372	Probe #19058 used
65		13	52.0	128	21	AAA82189	N. meningitidis pa
66		13	52.0	157	21	AAC12134	Human secreted pro
67		13	52.0	192	21	AAC11863	Human secreted pro
68		13	52.0	251	19	AAX12455	Human biallelic po
69		13	52.0	263	22	AAI57288	Probe #25974 used
70	c	13	52.0	299	21	AAF09435	Fusarium venenatum
71	c	13	52.0	300	21	AAA01371	Human colon cancer
72	c	13	52.0	300	22	AAH65466	C glutamicum codin
73		13	52.0	367	22	AAI26915	Probe #16848 for g
74		13	52.0	367	22	AAI55715	Probe #24401 used
75	c	13	52.0	384	18	AAV01888	Human OTK27 gene.
76	c	13	52.0	391	22	AAF65138	Novel human polynu
77	c	13	52.0	398	22	AAF66457	Novel human polynu
78	c	13	52.0	400	21	AAA56847	Mycoplasma hyopneu
79		13	52.0	408	22	AAS10988	Cone snail cDNA en
80	c	13	52.0	423	20	AAX29975	Fragment of human
81		13	52.0	437	20	AAX40715	Human secreted pro
82		13	52.0	450	22	AAI11054	Probe #987 for gen
83		13	52.0	450	22	AAI32317	Probe #1003 used t
84		13	52.0	450	22	AAI00974	Probe #965 used to

c	85	13	52.0	453	22	AAI37182	Probe #5868 used t
	86	13	52.0	458	22	AAI33177	Probe #1863 used t
	87	13	52.0	458	22	AAI52434	Probe #21120 used
c	88	13	52.0	509	21	AAC55960	Eucalyptus grandis
	89	13	52.0	543	22	AAI17715	Probe #7648 for ge
	90	13	52.0	543	22	AAI42652	Probe #11338 used
	91	13	52.0	560	22	AAI39351	Probe #8037 used t
	92	13	52.0	575	22	AAI18864	Probe #8797 for ge
	93	13	52.0	575	22	AAI43986	Probe #12672 used
	94	13	52.0	583	16	AAT05395	Septoria nodorum i
	95	13	52.0	583	19	AAV62502	Septoria nodorum P
	96	13	52.0	583	22	AAS08423	Internal transcrib
	97	13	52.0	596	22	AAI44078	Probe #12764 used
	98	13	52.0	642	22	AAF87980	E. coli orotate ph
c	99	13	52.0	666	21	AAF12304	Aspergillus oryzae
c	100	13	52.0	686	21	AAF12060	Aspergillus oryzae

ALIGNMENTS

```
RESULT   1
AAI62941
ID   AAI62941 standard; DNA; 839 BP.
XX
AC   AAI62941;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human genomic DNA SEQ ID NO 269.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant;  immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200155449-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01346.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUL-2000; 2000US-0216880.
PR   14-JUL-2000; 2000US-0218290.
PR   14-AUG-2000; 2000US-0225447.
PR   01-SEP-2000; 2000US-0229343.
PR   06-SEP-2000; 2000US-0230437.
PR   08-SEP-2000; 2000US-0231243.
PR   25-SEP-2000; 2000US-0234997.
PR   29-SEP-2000; 2000US-0236367.
PR   13-OCT-2000; 2000US-0239937.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246528.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249265.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-476225/51.
XX
PT   Novel plasma membrane associated proteins useful for diagnosing,
PT   treating, preventing and/or prognosing disorders related to the
PT   proteins, including cancer, immune response and neuronal disorders  -
XX
PS   Example 2; SEQ ID NO 269; 532pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (AAI62752-AAI62961) and proteins
CC   (AAM42347-AAM42415) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 839 BP; 205 A; 227 C; 243 G; 161 T; 3 other;


  Query Match             100.0%;  Score 25;  DB 22;  Length 839;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db      95 ggatctgctgtttgaagcgcagcag 119


RESULT   2
AAX01303
ID   AAX01303 standard; DNA; 1041 BP.
XX
AC   AAX01303;
XX
DT   12-APR-1999  (first entry)
XX
DE   E. coli biotin synthetase (BioB) coding sequence.
XX
KW   DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW   biotin synthase; biotin production; vitamin H; BioB; ss.
XX
OS   Escherichia coli.
XX
PN   US5869719-A.
XX
PD   09-FEB-1999.
XX
PF   30-APR-1997;   97US-0846338.
XX
PR   30-APR-1997;   97US-0846338.
PR   08-MAR-1995;   95US-0401068.
XX
```

```
PA   (NOVS ) NOVARTIS FINANCE CORP.
XX
PI   Patton DA;
XX
DR   WPI; 1999-152902/13.
DR   P-PSDB; AAW73906.
XX
PT   Transgenic plants with high biotin levels - transformed with DNA
PT   encoding di:amino-pelargonic acid amino-transferase or biotin
PT   synthase
XX
PS   Example 2; Column 37-40; 34pp; English.
XX
CC   This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC   be used in the transgenic plant of the invention. The transgenic plant,
CC   plant cell or plant tissue is transformed with a chimeric gene encoding
CC   diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC   produces more biotin than a non-transgenic plant, cell or tissue. The
CC   plant is used as an improved dietary source of biotin (vitamin H) for
CC   humans or animals.
XX
SQ   Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 20;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db      63 ggatctgctgtttgaagcgcagcag 87


RESULT   3
AAN91329
ID   AAN91329 standard; DNA; 1084 BP.
XX
AC   AAN91329;
XX
DT   15-FEB-1990  (first entry)
XX
DE   E.coli Bio B gene.
XX
KW   E.coli; Bio B gene; biotin.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             24..1064
FT                   /*tag=a
XX
PN   GB2216530-A.
XX
PD   11-OCT-1989.
XX
PF   17-MAR-1989;   89GB-0006210.
XX
PR   22-MAR-1988;   88GB-0006804.
PR   17-MAR-1989;   89GB-0006210.
XX
PA   (UKAG-) UK MIN. AGRIC. FISH.
XX
PI   Pearson BM,  McKee RA;
XX
DR   WPI; 1989-295085/41. P-PSDB P91392
XX
PT   Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT   - derived from E.coli and capable of replication and expression in other
PT   microorganisms, esp. yeast.
XX
PS   Table 3; page 33-4; 52pp; English.
XX
CC   The gene can be used in a plasmid for expression of enzymes of the biotin
CC   synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC   are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC   Lactobacillus. Insertion of bio B improves biotin yields in
CC   microorganisms which export biotin, or enables growth in media contg.
CC   little or no biotin of organisms unable to synthesise biotin for their
CC   own use.
XX
SQ   Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 10;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db      86 ggatctgctgtttgaagcgcagcag 110


RESULT   4
AAN60496
ID   AAN60496 standard; DNA; 1121 BP.
XX
AC   AAN60496;
XX
DT   17-OCT-1991  (first entry)
XX
DE   Sequence encoding biotin synthesising enzyme.
XX
KW   Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX
FH   Key             Location/Qualifiers
FT   CDS             42..1082
FT                   /*tag=  a
XX
PN   JP61149091-A.
XX
PD   07-JUL-1986.
XX
PF   24-DEC-1984;   84JP-0272605.
XX
PR   24-DEC-1984;   84JP-0272605.
XX
PA   (NIPS ) NIPPON SODA KK.
XX
DR   WPI; 1986-216622/33.
DR   P-PSDB; AAP60536.
XX
PT   Double stranded DNA encoding biotin synthesising enzyme -
PT   comprises transformed mutant E.coli strain contg. cyclic doubled
PT   stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX
PS   Disclosure; Page 534; 23pp; Japanese.
XX
CC   The sequence may be expressed by a transformed E.coli host, cultured
CC   in a medium containing desthiobiotin.
XX
SQ   Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 7;  Length 1121;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db     104 ggatctgctgtttgaagcgcagcag 128


RESULT   5
AAQ62386
ID   AAQ62386 standard; DNA; 5872 BP.
XX
```

```
AC   AAQ62386;
XX
DT   16-NOV-1994  (first entry)
XX
DE   Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX
KW   Biotin; expression; enterobacteria; vitamin H; synthesis;
KW   plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
KW   promoter ptac; biotin synthase; KAPA synthase;
KW   8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;
KW   dethiobiotin synthase; DAPA synthase;
KW   S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW   seborrhoea; dermatitis; ds.
XX
OS   Escherichia coli DSM498.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..96
FT                   /*tag= a
FT                   /function= "promoter ptac"
FT                   /evidence= EXPERIMENTAL
FT   -35_signal      23..28
FT                   /*tag= b
FT                   /standard_name- "promoter ptac"
FT   -10_signal      45..50
FT                   /*tag= c
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name- "promoter ptac"
FT   RBS             105..109
FT                   /*tag= d
FT                   /evidence= EXPERIMENTAL
FT                   /standard_name= "bioB RBS no. 9"
FT   CDS             117..1157
FT                   /*tag= e
FT                   /product= "biotin synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioB"
FT                   /number= 1
FT   RBS             1141..1146
FT                   /*tag= f
FT                   /standard_name= "bioF RBS"
FT   CDS             1154..2311
FT                   /*tag= g
FT                   /EC_number= 2.3.1.47
FT                   /product= "KAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioF"
FT                   /number= 2
FT                   /standard_name= "8-amino-7-oxononanoate synthase"
FT   RBS             2284..2288
FT                   /*tag= h
FT                   /standard_name= "bioC RBS"
FT   CDS             2295..3050
FT                   /*tag= i
FT                   /function= "involved in pimeloyl-CoA synthesis"
FT                   /product= "protein"
FT                   /gene= "bioC"
FT                   /number= 3
FT   RBS             3030..3033
FT                   /*tag= j
FT                   /standard_name= "bioD RBS"
FT   CDS             3043..3753
FT                   /*tag= k
FT                   /EC_number= 6.3.3.3
FT                   /product= "DTB synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioD15"
FT                   /number= 4
FT                   /standard_name= "dethiobiotin synthase"
FT   misc_RNA        3712..3750
FT                   /*tag= l
FT                   /note= "bioD15 substitution"
FT   RBS             3742..3746
FT                   /*tag= m
FT                   /standard_name= "bioA RBS"
FT   CDS             3750..5039
FT                   /*tag= n
FT                   /EC_number= 2.6.1.62
FT                   /product= "DAPA synthase"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "bioA"
FT                   /number= 5
FT                   /standard_name= "S-adenosyl-L-methionine: 8-amino-
FT                                   7-oxononanoate aminotransferase"
FT   RBS             5088..5093
FT                   /*tag= o
FT                   /standard_name= "ORFI RBS"
FT   CDS             5098..5574
FT                   /*tag= p
FT                   /function= "unknown, involved in biotin synthsesis"
FT                   /product= "protein"
FT                   /evidence= EXPERIMENTAL
FT                   /gene= "ORFI"
FT                   /number= 6
FT   terminator      5583..5644
FT                   /*tag= q
FT                   /standard_name= "rho-independent transcriptional
FT                                   terminator"
FT   stem_loop       5583..5605
FT                   /*tag= r
XX
PN   WO9408023-A.
XX
PD   14-APR-1994.
XX
PF   01-OCT-1993;    93WO-EP02688.
XX
PR   02-OCT-1992;    92CH-0003124.
PR   15-JUL-1993;    93CH-0002134.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Birch O,  Brass J,  Fuhrmann M,  Shaw N;
XX
DR   WPI; 1994-135587/16.
DR   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX
PT   Biotechnological biotin prodn. using enterobacterial biotin-gene
PT   - providing vitamin H in high yield
XX
PS   Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX
CC   The sequence is derived from plasmid pB030A-15/9 contg. the
CC   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
CC   of biotin, arranged in a transcription unit.  Microorganisms
CC   contg. these DNA fragments or plasmids may be used in the prodn.
CC   of biotin.  Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
CC   loss of appetite and tiredness.
XX
SQ   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;


  Query Match             100.0%;  Score 25;  DB 15;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db     179 ggatctgctgtttgaagcgcagcag 203


RESULT   6
AAX20559/c
ID   AAX20559 standard; DNA; 5504 BP.
XX
AC   AAX20559;
```

XX
DT 05-MAY-1999 (first entry)
XX
DE Polynucleotide sequence from the genome of Treponema pallidum.
XX
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
XX
OS Treponema pallidum.
XX
PN WO9859034-A2.
XX
PD 30-DEC-1998.
XX
PF 23-JUN-1998; 98WO-US13041.
XX
PR 24-JUN-1997; 97US-0050667.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Fraser CM;
XX
DR WPI; 1999-081273/07.
XX
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
XX
PS Claim 1; Page 519-522; 1150pp; English.
XX
CC AAX20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX
SQ Sequence 5504 BP; 1474 A; 1764 C; 1181 G; 1076 T; 9 other;

Query Match 68.0%; Score 17; DB 20; Length 5504;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 gctgtttgaagcgcagc 23
||||||||||||||||||
Db 1365 GCTGTTTGAAGCGCAGC 1349

RESULT 7
AAH65216/c
ID AAH65216 standard; DNA; 801 BP.
XX
AC AAH65216;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 251.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG89997.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 8; SEQ ID NO: 251; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 801 BP; 184 A; 221 C; 215 G; 181 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 801;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 ctgtttgaagcgcagc 23
||||||||||||||||
Db 784 CTGTTTGAAGCGCAGC 769

RESULT 8
AAF71927/c
ID AAF71927 standard; DNA; 924 BP.
XX
AC AAF71927;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:349.
XX
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100843-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00923.
XX
PR 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-137957/14.
DR P-PSDB; AAB79808.
XX
PT Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX
PS Claim 3; Page 676-677; 1737pp; English.
XX
CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
SQ Sequence 924 BP; 214 A; 252 C; 239 G; 219 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 924;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 ctgtttgaagcgcagc 23
||||||||||||||||
Db 884 CTGTTTGAAGCGCAGC 869

RESULT 9
AAH64966/c
ID AAH64966 standard; DNA; 349980 BP.
XX
AC AAH64966;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 1.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT -
XX
PS Claim 7; SEQ ID NO: 1; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 349980 BP; 79703 A; 91547 C; 98381 G; 80349 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 349980;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       8 ctgtttgaagcgcagc 23
           ||||||||||||||||
Db 238128 CTGTTTGAAGCGCAGC 238113
```

RESULT 10
AAH01365
ID AAH01365 standard; DNA; 810 BP.
XX
AC AAH01365;
XX
DT 24-JUL-2001 (first entry)
XX
DE Serratia marcescens nucleotide sequence SEQ ID NO:1356.
XX
KW Species specific; genus specific; family specific; probe; detection;
KW identification; algal; archaeal; bacterial; fungal; parasitical;
KW microorganism; diagnosis; translation elongation factor Tu; toxin;
KW translation elongation factor G; RecA recombinase; resistance;
KW catalytic subunit of proton-translocating ATPase; antimicrobial;
KW vaccine; primer; ds.
XX
OS Serratia marcescens.
XX
PN WO200123604-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-CA01150.
XX
PR 28-SEP-1999; 99CA-2283458.
PR 19-MAY-2000; 2000CA-2307010.
XX
PA (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
XX
PI Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;
PI Picard FJ, Roy PH;
XX
DR WPI; 2001-245006/25.
XX
PT Nucleic acid sequences are used to generate universal probes and
PT primers which can be used to identify and detect the presence of algal,
PT archaeal, bacterial, fungal and parasitical species in a test sample -
XX
PS Disclosure; Page 1131-1132; 1580pp; English.
XX
CC The present invention describes a method for generating a repertory of
CC nucleic acids of tuf, fus, atpD and/or recA genes from which probes
CC and/or primers are derived. The method comprises amplifying the nucleic
CC acids of determined algal, archaeal, bacterial, fungal and parasitical
CC species with a combination of defined primer pairs. The method can be
CC used for producing probes and/or primers for detecting one or more
CC related microorganisms e.g. algae, archaea, bacteria, fungi and
CC parasites, for universal detection and for specific and ubiquitous
CC detection and identification of an algal, archaeal, bacterial, fungal
CC and parasitical species, genus, family and group. A nucleic acid (I)
CC obtained using the method of the invention can be used for the universal
CC detection of any bacterium, fungus or parasite in a sample and for the
CC detection of at least one antimicrobial agent resistance gene or at
CC least one toxin gene. hexA nucleic acids are used for the specific and
CC ubiquitous detection and for identification of Streptococcus pneumoniae.
CC (I) can be used to design a therapeutic agent which is effective against
CC microorganisms. Microbial species or genus or family or phylum or group
CC which can be detected include Abiotrophia adiacens, Bordetella sp.,
CC Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
CC Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,
CC Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
CC provides faster results than substrate specificity tests as results can
CC be determined in an hour and improved accuracy is also achieved.
CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
CC which are given in the exemplification of the present invention.
XX
SQ Sequence 810 BP; 129 A; 270 C; 268 G; 143 T; 0 other;

Query Match 60.0%; Score 15; DB 22; Length 810;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       8 ctgtttgaagcgcag 22
           |||||||||||||||
Db     742 ctgtttgaagcgcag 756
```

RESULT 11
AAZ23024/c
ID AAZ23024 standard; cDNA; 1689 BP.
XX
AC AAZ23024;
XX
DT 17-JAN-2000 (first entry)
XX
DE Rat kd312 polypeptide encoding cDNA.
XX
KW kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW heart attack; head trauma; neurodegenerative disease; rat;
KW Parkinson's disease; Alzheimer's disease; ss.
XX
OS Rattus sp.
XX
PN WO9950288-A2.
XX
PD 07-OCT-1999.
XX
PF 30-MAR-1999; 99WO-US06993.
XX
PR 31-MAR-1998; 98US-0053374.
XX
PA (AMGE-) AMGEN INC.
XX
PI Yen K;
XX
DR WPI; 1999-601322/51.
DR P-PSDB; AAY42694.
XX
PT kd312 polypeptides useful for treating diseases and disorders
PT associated with alterations in cell proliferation and cell death -
XX
PS Claim 2; Fig 9; 85pp; English.
XX
CC The invention provides nucleic acid molecules encoding human and rat
CC kd312 polypeptides. The kd312 polypeptides can be expressed by standard
CC recombinant methodology. The kd312 sequences, and the antibodies against
CC the proteins may be used to treat or diagnose the presence or progression
CC of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
CC stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
CC Parkinson's disease and Alzheimer's disease). The present sequence
CC represents the rat kd312 cDNA sequence.
XX
SQ Sequence 1689 BP; 417 A; 462 C; 434 G; 376 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 1689;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ggatctgctgtttga 15
           |||||||||||||||
Db     495 GGATCTGCTGTTTGA 481
```

RESULT 12
AAC44132/c
ID AAC44132 standard; DNA; 2022 BP.
XX
AC AAC44132;

XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 41750.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

```
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             60.0%;  Score 15;  DB 21;  Length 2022;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaagc 18
           |||||||||||||||
Db    1268 TCTGCTGTTTGAAGC 1254


RESULT  13
AAZ23025/c
ID   AAZ23025 standard; DNA; 3079 BP.
XX
AC   AAZ23025;
XX
DT   17-JAN-2000  (first entry)
XX
DE   Rat kd312 genomic DNA sequence.
XX
KW   kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW   heart attack; head trauma; neurodegenerative disease; rat;
KW   Parkinson's disease; Alzheimer's disease; ss.
XX
OS   Rattus sp.
XX
PN   WO9950288-A2.
XX
PD   07-OCT-1999.
XX
PF   30-MAR-1999;   99WO-US06993.
XX
PR   31-MAR-1998;   98US-0053374.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Yen K;
XX
DR   WPI; 1999-601322/51.
DR   P-PSDB; AAY42694.
XX
PT   kd312 polypeptides useful for treating diseases and disorders
PT   associated with alterations in cell proliferation and cell death
XX
PS   Claim 2; Fig 7; 85pp; English.
XX
CC   The invention provides nucleic acid molecules encoding human and rat
CC   kd312 polypeptides. The kd312 polypeptides can be expressed by standard
CC   recombinant methodology. The kd312 sequences, and the antibodies against
CC   the proteins may be used to treat or diagnose the presence or progression
CC   of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
CC   stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
CC   Parkinson's disease and Alzheimer's disease). The present sequence
CC   represents the rat kd312 genomic DNA sequence.
XX
SQ   Sequence 3079 BP; 708 A; 843 C; 829 G; 699 T; 0 other;


  Query Match             60.0%;  Score 15;  DB 20;  Length 3079;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttga 15
           |||||||||||||||
Db     803 GGATCTGCTGTTTGA 789


RESULT  14
AAF65487/c
ID   AAF65487 standard; cDNA; 336 BP.
XX
AC   AAF65487;
XX
DT   09-APR-2001  (first entry)
XX
DE   Novel human polynucleotide, SEQ ID NO: 1243.
XX
KW   Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW   breast cancer; lung cancer; cancer detection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102568-A2.
XX
PD   11-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18374.
XX
PR   02-JUL-1999;   99US-0142310.
PR   02-JUL-1999;   99US-0142311.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Klinger J,  Kassam A;
```

PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitiz D;
PI Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -
XX
PS Claim 9; Page 722; 1046pp; English.
XX
CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.
XX
SQ Sequence 336 BP; 97 A; 84 C; 59 G; 96 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 336;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 atctgctgtttgaa 16
||||||||||||||
Db 159 ATCTGCTGTTTGAA 146

RESULT 15
AAV87319/c
ID AAV87319 standard; cDNA; 379 BP.
XX
AC AAV87319;
XX
DT 27-APR-1999 (first entry)
XX
DE EST clone BP797.
XX
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS Homo sapiens.
XX
PN WO9845435-A2.
XX
PD 15-OCT-1998.
XX
PF 10-APR-1998; 98WO-US06954.
XX
PR 10-APR-1997; 97US-0835913.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-070076/06.
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX
PS Claim 1; Page 527; 633pp; English.
XX
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
XX
SQ Sequence 379 BP; 124 A; 68 C; 70 G; 116 T; 1 other;

Query Match 56.0%; Score 14; DB 20; Length 379;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gatctgctgtttga 15
||||||||||||||
Db 129 GATCTGCTGTTTGA 116

RESULT 16
AAH10789
ID AAH10789 standard; cDNA; 393 BP.
XX
AC AAH10789;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (3'-primer) SEQ ID NO:7624.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3; SEQ ID 7624; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602

```
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 393 BP; 89 A; 100 C; 115 G; 82 T; 7 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 393;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db     293 tctgctgtttgaag 306


RESULT  17
AAI12900
ID   AAI12900 standard; DNA; 469 BP.
XX
AC   AAI12900;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #2833 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 2833; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 469 BP; 81 A; 90 C; 145 G; 153 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 469;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 gatctgctgtttga 15
           ||||||||||||||
Db     295 gatctgctgtttga 308


RESULT  18
AAI34258
ID   AAI34258 standard; DNA; 469 BP.
XX
AC   AAI34258;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #2944 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 2944; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 469 BP; 81 A; 90 C; 145 G; 153 T; 0 other;
```

```
  Query Match            56.0%;  Score 14;  DB 22;  Length 469;
  Best Local Similarity  100.0%;  Pred. No. 63;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 gatctgctgtttga 15
          ||||||||||||||
Db    295 gatctgctgtttga 308


RESULT  19
AAI02819
ID   AAI02819 standard; DNA; 469 BP.
XX
AC   AAI02819;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Probe #2810 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
KW   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PF   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
PT   in a human breast -
XX
PS   Claim 25; SEQ ID No 2810; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
CC   The present sequence is one such probe. The probes are useful for
CC   measuring human gene expression in a human breast sample, where the probe
CC   hybridises at high stringency to a nucleic acid expressed in the human
CC   breast. The probes are useful for predicting, diagnosing, grading,
CC   staging, monitoring and prognosing diseases of the human breast,
CC   particularly those diseases with polygenic aetiology. The diseases
CC   include: breast cancer, disorders of development, inflammatory diseases
CC   of the breast, fibrocystic changes, proliferative breast disease and
CC   non-carcinoma tumours.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 469 BP; 81 A; 90 C; 145 G; 153 T; 0 other;


  Query Match            56.0%;  Score 14;  DB 22;  Length 469;
  Best Local Similarity  100.0%;  Pred. No. 63;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 gatctgctgtttga 15
          ||||||||||||||
Db    295 gatctgctgtttga 308


RESULT  20
AAI22139
ID   AAI22139 standard; DNA; 476 BP.
XX
AC   AAI22139;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #12072 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 25; SEQ ID No 12072; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging
CC   of diseases of the cervix, notably cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;


  Query Match            56.0%;  Score 14;  DB 22;  Length 476;
  Best Local Similarity  100.0%;  Pred. No. 63;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 gatctgctgtttga 15
          ||||||||||||||
Db    135 gatctgctgtttga 148


RESULT  21
AAI47435
ID   AAI47435 standard; DNA; 476 BP.
XX
AC   AAI47435;
XX
DT   17-OCT-2001  (first entry)
XX
```

DE Probe #16121 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 16121; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;

```
  Query Match             56.0%;  Score 14;  DB 22;  Length 476;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 gatctgctgtttga 15
           ||||||||||||||
Db     135 gatctgctgtttga 148
```

RESULT 22
AAI07839
ID AAI07839 standard; DNA; 476 BP.
XX
AC AAI07839;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #7830 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US00661.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 25; SEQ ID No 7830; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 476 BP; 84 A; 92 C; 142 G; 158 T; 0 other;

```
  Query Match             56.0%;  Score 14;  DB 22;  Length 476;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 gatctgctgtttga 15
           ||||||||||||||
Db     135 gatctgctgtttga 148
```

RESULT 23
AAA44124/c
ID AAA44124 standard; cDNA; 619 BP.
XX
AC AAA44124;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:699.
XX
KW Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200021991-A1.
XX
PD 20-APR-2000.

XX
PF 15-OCT-1999; 99WO-US24206.
XX
PR 15-OCT-1998; 98US-0104436.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR;
XX
DR WPI; 2000-317938/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (sESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 383; 803pp; English.
XX
CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (sESTs), isolated from human, mouse, chicken and rat
CC tissue sources. The sESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnerary; antiulcer; osteopathic; neuroprotective;
CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The sESTs can be used for gene
CC therapy and in vaccines. The sESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 619 BP; 193 A; 123 C; 125 G; 178 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 619;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 atctgctgtttgaa 16
||||||||||||||
Db 43 ATCTGCTGTTTGAA 30

RESULT 24
AAZ00347/c
ID AAZ00347 standard; DNA; 753 BP.
XX
AC AAZ00347;
XX
DT 26-OCT-1999 (first entry)
XX
DE Nucleotide sequence of human hsFATP3.
XX
KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO9936537-A2.
XX
PD 22-JUL-1999.
XX
PF 14-JAN-1999; 99WO-US00182.
XX
PR 14-JAN-1999; 99US-0232201.
PR 15-JAN-1998; 98US-0071374.
PR 20-JUL-1998; 98US-0093491.
PR 04-DEC-1998; 98US-0110941.
PR 14-JAN-1999; 99US-0232195.
PR 14-JAN-1999; 99US-0232197.
PR 14-JAN-1999; 99US-0232200.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
DR WPI; 1999-444398/37.
DR P-PSDB; AAY14937.
XX
PT Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
PS Example 1; Fig 16; 255pp; English.
XX
CC The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
SQ Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 753;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tctgctgtttgaag 17
||||||||||||||
Db 436 TCTGCTGTTTGAAG 423

RESULT 25
AAF89005/c
ID AAF89005 standard; DNA; 753 BP.
XX
AC AAF89005;
XX
DT 06-JUL-2001 (first entry)
XX
DE Human FATP3 coding sequence SEQ ID NO: 14.
XX
KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
KW weight control; tuberculosis; TB; anti-fungal; ds.
XX
OS Homo sapiens.
XX
PN WO200121795-A2.
XX
PD 29-MAR-2001.
XX

```
PF   21-SEP-2000; 2000WO-US25891.
XX
PR   23-SEP-1999;   99US-0405504.
PR   23-SEP-1999;   99US-0405505.
PR   16-DEC-1999;   99US-0465280.
PR   17-FEB-2000; 2000US-0506252.
PR   06-JUL-2000; 2000US-0611197.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Stahl A,  Hirsch DJ,  Lodish HF,  Gimeno RE,  Tartaglia LA;
XX
DR   WPI; 2001-354783/37.
DR   P-PSDB; AAB83227.
XX
PT   New fatty acid transport proteins (FATPs) useful for the manufacture of
PT   medicament for treating obesity, diabetes and heart disease -
XX
PS   Disclosure; Fig 16; 287pp; English.
XX
CC   The present invention provides the protein and coding sequences of fatty
CC   acid transport proteins (FATPs) from a number of species, including
CC   FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
CC   from the mouse, FATPa and b from C. elegans, and FATP from Aspergillus
CC   nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
CC   tuberculosis and Cochliobolus heterostrophus. The FATP from M.
CC   tuberculosis can be used to identify inhibitors which can then be used to
CC   treat TB. That from M. grisea (also known as rice blast fungus) can be
CC   used to develop anti-fungal agents capable of preventing infection of
CC   rice. Those from the human can be used to develop treatments for
CC   diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
CC   present sequence is one of the sequences described in the exemplification
CC   of the invention.
XX
SQ   Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 63;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db     436 TCTGCTGTTTGAAG 423


RESULT  26
AAZ00358/c
ID   AAZ00358 standard; DNA; 1173 BP.
XX
AC   AAZ00358;
XX
DT   26-OCT-1999  (first entry)
XX
DE   Nucleotide sequence of human hsFATP3.
XX
KW   Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW   fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO9936537-A2.
XX
PD   22-JUL-1999.
XX
PF   14-JAN-1999;   99WO-US00182.
XX
PR   14-JAN-1999;   99US-0232201.
PR   15-JAN-1998;   98US-0071374.
PR   20-JUL-1998;   98US-0093491.
PR   04-DEC-1998;   98US-0110941.
PR   14-JAN-1999;   99US-0232195.
PR   14-JAN-1999;   99US-0232197.
PR   14-JAN-1999;   99US-0232200.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Gimeno RE,  Hirsch DJ,  Lodish HF,  Stahl A,  Tartaglia LA;
XX
DR   WPI; 1999-444398/37.
DR   P-PSDB; AAY14948.
XX
PT   Fatty acid transport proteins and related polynucleotides, useful
PT   for treating obesity, diabetes and heart disease
XX
PS   Claim 31; Fig 48; 255pp; English.
XX
CC   The invention provides a family of fatty acid transport proteins (FATPs)
CC   that mediate transport of long chain fatty acids (LCFAs) across cell
CC   membranes into cells. Human and murine FATP proteins and nucleic acids
CC   encoding the proteins are provided. The FATP proteins can be produced
CC   by standard recombinant methodology. Fatty acid uptake by cells can be
CC   modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC   In particular, antisense oligonucleotides can be used to modulate FATP
CC   biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC   uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC   muscle or liver by administration of a complex of the agent and a FATP6
CC   binding moiety. DNA encoding FATP proteins can be used as a reference
CC   used in detecting variant alleles or homologues. Altering the LCFA uptake
CC   by administering an inhibitor or enhancer of FATP transport function in
CC   the small intestine can decrease or increase calories available as fats,
CC   and can decrease or increase circulating fatty acids. Blocking the
CC   function of FATP4 and also FATP2, is useful for treating obesity,
CC   diabetes and heart disease.
XX
SQ   Sequence 1173 BP; 247 A; 327 C; 351 G; 248 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 1173;
  Best Local Similarity   100.0%;  Pred. No. 64;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db     853 TCTGCTGTTTGAAG 840


RESULT  27
AAF89020/c
ID   AAF89020 standard; DNA; 1173 BP.
XX
AC   AAF89020;
XX
DT   06-JUL-2001  (first entry)
XX
DE   Human FATP3 partial coding sequence SEQ ID NO: 50.
XX
KW   Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
KW   yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
KW   weight control; tuberculosis; TB; anti-fungal; ds.
XX
OS   Homo sapiens.
XX
PN   WO200121795-A2.
XX
PD   29-MAR-2001.
XX
PF   21-SEP-2000; 2000WO-US25891.
XX
PR   23-SEP-1999;   99US-0405504.
PR   23-SEP-1999;   99US-0405505.
PR   16-DEC-1999;   99US-0465280.
PR   17-FEB-2000; 2000US-0506252.
PR   06-JUL-2000; 2000US-0611197.
```

```
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Stahl A,  Hirsch DJ,  Lodish HF,  Gimeno RE,  Tartaglia LA;
XX
DR   WPI; 2001-354783/37.
DR   P-PSDB; AAB83248.
XX
PT   New fatty acid transport proteins (FATPs) useful for the manufacture of
PT   medicament for treating obesity, diabetes and heart disease -
XX
PS   Disclosure; Fig 48; 287pp; English.
XX
CC   The present invention provides the protein and coding sequences of fatty
CC   acid transport proteins (FATPs) from a number of species, including
CC   FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
CC   from the mouse, FATPa and b from C. elegans, and FATP from Aspergillus
CC   nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
CC   tuberculosis and Cochliobolus heterostrophus. The FATP from M.
CC   tuberculosis can be used to identify inhibitors which can then be used to
CC   treat TB. That from M. grisea (also known as rice blast fungus) can be
CC   used to develop anti-fungal agents capable of preventing infection of
CC   rice. Those from the human can be used to develop treatments for
CC   diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
CC   present sequence is one of the sequences described in the exemplification
CC   of the invention.
XX
SQ   Sequence 1173 BP; 247 A; 327 C; 351 G; 248 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 1173;
  Best Local Similarity   100.0%;  Pred. No. 64;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db    853 TCTGCTGTTTGAAG 840


RESULT  28
AAA50896/c
ID   AAA50896 standard; cDNA; 1244 BP.
XX
AC   AAA50896;
XX
DT   28-SEP-2000  (first entry)
XX
DE   Soybean PI metabolism protein clone, ssm.pk0068.c5 partial cDNA.
XX
KW   Enzyme; phosphatidylinositol-4-phosphate kinase; PI;
KW   phosphatidylinositol metabolism; signal transduction; membrane anchor;
KW   membrane trafficking pathway regulator; vacuole homeostasis regulator;
KW   transgenic plant; soybean; EST; expressed sequence tag; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..991
FT                   /*tag=  a
FT                   /partial
FT                   /product= "ssm.pk0068.c5"
XX
PN   WO200036119-A2.
XX
PD   22-JUN-2000.
XX
PF   17-DEC-1999;   99WO-US30182.
XX
PR   18-DEC-1998;   98US-0112925.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM,  Kinney AJ,  Miao G,  Rafalski JA,  Sakai H,  Weng Z;
XX
DR   WPI; 2000-431597/37.
DR   P-PSDB; AAY99751.
XX
PT   New isolated polynucleotide encoding phosphatidylinositol metabolism
PT   protein is useful for producing transgenic plants with an altered level
PT   of phosphatidylinositol -
XX
PS   Claim 2; Page 50-51; 65pp; English.
XX
CC   Phosphatidylinositol, PI, may be modified in cells to have roles in
CC   signal transduction, membrane associated protein anchoring and regulation
CC   of membrane trafficking pathways. The enzyme phosphatidylinositol-4-
CC   phosphate 5-kinase, converts phosphatidylinositol-4-phosphate to
CC   phosphatidylinositol-4,5-diphosphate. Phosphatidylinositol-4,5-
CC   diphosphate is thought to play a role in vacuole homeostasis regulation.
CC   The present sequence is the partial coding sequence of the soybean PI
CC   metabolism protein clone, ssm.pk0068.c5, isolated from an expressed
CC   sequence tag, EST. The present sequence may be used to create transgenic
CC   plants which express phosphatidylinositol-4-phosphate kinase, and hence
CC   be useful for modifying plant phosphatidylinositols.
XX
SQ   Sequence 1244 BP; 392 A; 207 C; 274 G; 371 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 21;  Length 1244;
  Best Local Similarity   100.0%;  Pred. No. 64;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gtttgaagcgcagc 23
           ||||||||||||||
Db     886 GTTTGAAGCGCAGC 873


RESULT  29
AAQ04684
ID   AAQ04684 standard; DNA; 1536 BP.
XX
AC   AAQ04684;
XX
DT   05-OCT-1990  (first entry)
XX
DE   SalI/KpnI fragment of pUO1 contg. haloacetate dehalogenase H-1.
XX
KW   Haloacetate dehalogenase; H-1; H-2; ss.
XX
OS   Moraxell sp. B.
XX
FH   Key             Location/Qualifiers
FT   CDS             413..1296
FT                   /*tag=  a
FT                   /label=H-1_gene
XX
PN   JP02109981-A.
XX
PD   23-APR-1990.
XX
PF   08-APR-1988;   88JP-0085393.
XX
PR   08-APR-1988;   88JP-0085393.
XX
PA   (MITN ) MITSUBISHI GAS CHEM KK.
XX
DR   WPI; 1990-168356/22.
DR   P-PSDB; AAR06435.
XX
PT   Haloacetate dehalogenase coding gene - used to raise
PT   decomposition activity of arganic halogenated cpds.
XX
PS   Claim 1; Fig 2; 8pp; Japanese.
XX
CC   The DNA can be ligated into a vector which can then be used to
```

CC transform E.coli for the prodn. of H-1. The protein is used in
CC the decomposition of organic halogenated cpds.
CC See also AAQ04685.
XX
SQ Sequence 1536 BP; 315 A; 445 C; 449 G; 326 T; 1 other;

Query Match 56.0%; Score 14; DB 11; Length 1536;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 ttgaagcgcagcag 25
||||||||||||||
Db 320 ttgaagcgcagcag 333

RESULT 30
AAI58178/c
ID AAI58178 standard; cDNA; 1699 BP.
XX
AC AAI58178;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 381.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM39022.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 381; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1699 BP; 340 A; 490 C; 499 G; 370 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 1699;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tctgctgtttgaag 17
||||||||||||||
Db 1393 TCTGCTGTTTGAAG 1380

RESULT 31
AAZ00367/c
ID AAZ00367 standard; DNA; 1998 BP.
XX
AC AAZ00367;
XX
DT 26-OCT-1999 (first entry)
XX
DE Nucleotide sequence of murine mmFATP3.
XX
KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS Mus sp.
XX
PN WO9936537-A2.
XX
PD 22-JUL-1999.
XX
PF 14-JAN-1999; 99WO-US00182.
XX
PR 14-JAN-1999; 99US-0232201.
PR 15-JAN-1998; 98US-0071374.
PR 20-JUL-1998; 98US-0093491.
PR 04-DEC-1998; 98US-0110941.
PR 14-JAN-1999; 99US-0232195.
PR 14-JAN-1999; 99US-0232197.
PR 14-JAN-1999; 99US-0232200.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
DR WPI; 1999-444398/37.
DR P-PSDB; AAY14957.
XX
PT Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
PS Example 1; Fig 66; 255pp; English.
XX
CC The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference

```
CC   used in detecting variant alleles or homologues. Altering the LCFA uptake
CC   by administering an inhibitor or enhancer of FATP transport function in
CC   the small intestine can decrease or increase calories available as fats,
CC   and can decrease or increase circulating fatty acids. Blocking the
CC   function of FATP4 and also FATP2, is useful for treating obesity,
CC   diabetes and heart disease.
XX
SQ   Sequence 1998 BP; 422 A; 567 C; 597 G; 412 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 20;  Length 1998;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db   1690 TCTGCTGTTTGAAG 1677


RESULT  32
AAF89029/c
ID   AAF89029 standard; DNA; 1998 BP.
XX
AC   AAF89029;
XX
DT   06-JUL-2001  (first entry)
XX
DE   Murine FATP3 partial coding sequence SEQ ID NO: 68.
XX
KW   Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
KW   yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
KW   weight control; tuberculosis; TB; anti-fungal; ds.
XX
OS   Mus musculus.
XX
PN   WO200121795-A2.
XX
PD   29-MAR-2001.
XX
PF   21-SEP-2000; 2000WO-US25891.
XX
PR   23-SEP-1999;   99US-0405504.
PR   23-SEP-1999;   99US-0405505.
PR   16-DEC-1999;   99US-0465280.
PR   17-FEB-2000; 2000US-0506252.
PR   06-JUL-2000; 2000US-0611197.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Stahl A,  Hirsch DJ,  Lodish HF,  Gimeno RE,  Tartaglia LA;
XX
DR   WPI; 2001-354783/37.
DR   P-PSDB; AAB83257.
XX
PT   New fatty acid transport proteins (FATPs) useful for the manufacture of
PT   medicament for treating obesity, diabetes and heart disease -
XX
PS   Disclosure; Fig 66; 287pp; English.
XX
CC   The present invention provides the protein and coding sequences of fatty
CC   acid transport proteins (FATPs) from a number of species, including
CC   FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
CC   from the mouse, FATPa and b from C. elegans, and FATP from Aspergillus
CC   nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
CC   tuberculosis and Cochliobolus heterostrophus. The FATP from M.
CC   tuberculosis can be used to identify inhibitors which can then be used to
CC   treat TB. That from M. grisea (also known as rice blast fungus) can be
CC   used to develop anti-fungal agents capable of preventing infection of
CC   rice. Those from the human can be used to develop treatments for
CC   diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
CC   present sequence is one of the sequences described in the exemplification
CC   of the invention.
XX
SQ   Sequence 1998 BP; 422 A; 567 C; 597 G; 412 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 1998;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db   1690 TCTGCTGTTTGAAG 1677


RESULT  33
AAZ00343/c
ID   AAZ00343 standard; DNA; 2087 BP.
XX
AC   AAZ00343;
XX
DT   26-OCT-1999  (first entry)
XX
DE   Nucleotide sequence of murine mmFATP3.
XX
KW   Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
KW   fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS   Mus musculus.
XX
PN   WO9936537-A2.
XX
PD   22-JUL-1999.
XX
PF   14-JAN-1999;   99WO-US00182.
XX
PR   14-JAN-1999;   99US-0232201.
PR   15-JAN-1998;   98US-0071374.
PR   20-JUL-1998;   98US-0093491.
PR   04-DEC-1998;   98US-0110941.
PR   14-JAN-1999;   99US-0232195.
PR   14-JAN-1999;   99US-0232197.
PR   14-JAN-1999;   99US-0232200.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Gimeno RE,  Hirsch DJ,  Lodish HF,  Stahl A,  Tartaglia LA;
XX
DR   WPI; 1999-444398/37.
DR   P-PSDB; AAY14933.
XX
PT   Fatty acid transport proteins and related polynucleotides, useful
PT   for treating obesity, diabetes and heart disease
XX
PS   Example 1; Fig 8; 255pp; English.
XX
CC   The invention provides a family of fatty acid transport proteins (FATPs)
CC   that mediate transport of long chain fatty acids (LCFAs) across cell
CC   membranes into cells. Human and murine FATP proteins and nucleic acids
CC   encoding the proteins are provided. The FATP proteins can be produced
CC   by standard recombinant methodology. Fatty acid uptake by cells can be
CC   modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC   In particular, antisense oligonucleotides can be used to modulate FATP
CC   biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC   uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC   muscle or liver by administration of a complex of the agent and a FATP6
CC   binding moiety. DNA encoding FATP proteins can be used as a reference
CC   used in detecting variant alleles or homologues. Altering the LCFA uptake
CC   by administering an inhibitor or enhancer of FATP transport function in
CC   the small intestine can decrease or increase calories available as fats,
CC   and can decrease or increase circulating fatty acids. Blocking the
CC   function of FATP4 and also FATP2, is useful for treating obesity,
CC   diabetes and heart disease.
XX
```

SQ Sequence 2087 BP; 441 A; 596 C; 624 G; 426 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2087;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db    1779 TCTGCTGTTTGAAG 1766
```

RESULT 34
AAF89001/c
ID AAF89001 standard; DNA; 2087 BP.
XX
AC AAF89001;
XX
DT 06-JUL-2001 (first entry)
XX
DE Murine FATP3 coding sequence SEQ ID NO: 6.
XX
KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
KW weight control; tuberculosis; TB; anti-fungal; ds.
XX
OS Mus musculus.
XX
PN WO200121795-A2.
XX
PD 29-MAR-2001.
XX
PF 21-SEP-2000; 2000WO-US25891.
XX
PR 23-SEP-1999; 99US-0405504.
PR 23-SEP-1999; 99US-0405505.
PR 16-DEC-1999; 99US-0465280.
PR 17-FEB-2000; 2000US-0506252.
PR 06-JUL-2000; 2000US-0611197.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;
XX
DR WPI; 2001-354783/37.
DR P-PSDB; AAB83223.
XX
PT New fatty acid transport proteins (FATPs) useful for the manufacture of
PT medicament for treating obesity, diabetes and heart disease -
XX
PS Examples; Fig 8; 287pp; English.
XX
CC The present invention provides the protein and coding sequences of fatty
CC acid transport proteins (FATPs) from a number of species, including
CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
CC from the mouse, FATPa and b from C. elegans, and FATP from Aspergillus
CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
CC tuberculosis can be used to identify inhibitors which can then be used to
CC treat TB. That from M. grisea (also known as rice blast fungus) can be
CC used to develop anti-fungal agents capable of preventing infection of
CC rice. Those from the human can be used to develop treatments for
CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
CC present sequence is one of the sequences described in the exemplification
CC of the invention.
XX
SQ Sequence 2087 BP; 441 A; 596 C; 624 G; 426 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2087;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db    1779 TCTGCTGTTTGAAG 1766
```

RESULT 35
AAZ00379/c
ID AAZ00379 standard; DNA; 2166 BP.
XX
AC AAZ00379;
XX
DT 26-OCT-1999 (first entry)
XX
DE Nucleotide sequence of human hsFATP5 gene.
XX
KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO9936537-A2.
XX
PD 22-JUL-1999.
XX
PF 14-JAN-1999; 99WO-US00182.
XX
PR 14-JAN-1999; 99US-0232201.
PR 15-JAN-1998; 98US-0071374.
PR 20-JUL-1998; 98US-0093491.
PR 04-DEC-1998; 98US-0110941.
PR 14-JAN-1999; 99US-0232195.
PR 14-JAN-1999; 99US-0232197.
PR 14-JAN-1999; 99US-0232200.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
DR WPI; 1999-444398/37.
DR P-PSDB; AAY14969.
XX
PT Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
PS Claim 31; Fig 94A-B; 255pp; English.
XX
CC The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
SQ Sequence 2166 BP; 370 A; 682 C; 695 G; 419 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2166;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db   1987 TCTGCTGTTTGAAG 1974


RESULT  36
AAF89041/c
ID   AAF89041 standard; DNA; 2166 BP.
XX
AC   AAF89041;
XX
DT   06-JUL-2001  (first entry)
XX
DE   Human FATP3 coding sequence SEQ ID NO: 101.
XX
KW   Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
KW   yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
KW   weight control; tuberculosis; TB; anti-fungal; ds.
XX
OS   Homo sapiens.
XX
PN   WO200121795-A2.
XX
PD   29-MAR-2001.
XX
PF   21-SEP-2000; 2000WO-US25891.
XX
PR   23-SEP-1999;   99US-0405504.
PR   23-SEP-1999;   99US-0405505.
PR   16-DEC-1999;   99US-0465280.
PR   17-FEB-2000; 2000US-0506252.
PR   06-JUL-2000; 2000US-0611197.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Stahl A,  Hirsch DJ,  Lodish HF,  Gimeno RE,  Tartaglia LA;
XX
DR   WPI; 2001-354783/37.
DR   P-PSDB; AAB83278.
XX
PT   New fatty acid transport proteins (FATPs) useful for the manufacture of
PT   medicament for treating obesity, diabetes and heart disease -
XX
PS   Disclosure; Fig 94; 287pp; English.
XX
CC   The present invention provides the protein and coding sequences of fatty
CC   acid transport proteins (FATPs) from a number of species, including
CC   FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
CC   from the mouse, FATPa and b from C. elegans, and FATP from Aspergillus
CC   nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
CC   tuberculosis and Cochliobolus heterostrophus. The FATP from M.
CC   tuberculosis can be used to identify inhibitors which can then be used to
CC   treat TB. That from M. grisea (also known as rice blast fungus) can be
CC   used to develop anti-fungal agents capable of preventing infection of
CC   rice. Those from the human can be used to develop treatments for
CC   diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
CC   present sequence is one of the sequences described in the exemplification
CC   of the invention.
XX
SQ   Sequence 2166 BP; 370 A; 682 C; 695 G; 419 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 2166;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db   1987 TCTGCTGTTTGAAG 1974


RESULT  37
AAH17621/c
ID   AAH17621 standard; cDNA; 2314 BP.
XX
AC   AAH17621;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:17145.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 17145; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2314 BP; 402 A; 704 C; 747 G; 461 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 2314;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db   2022 TCTGCTGTTTGAAG 2009
```

```
RESULT  38
AAH19223/c
ID   AAH19223 standard; cDNA; 2319 BP.
XX
AC   AAH19223;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 17 cDNA clone HSLCA15, SEQ ID NO:64.
XX
KW   Human; secreted protein; proliferative disorder; cancer;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder;  pregnancy-related disorder; tumour;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
PN   WO200132910-A2.
XX
PD   10-MAY-2001.
XX
PF   25-OCT-2000; 2000WO-US29362.
XX
PR   29-OCT-1999;   99US-0162240.
PR   30-JUN-2000; 2000US-0215131.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Birse CE,  Ni J,  Soppet DR;
XX
DR   WPI; 2001-335835/35.
DR   N-PSDB; AAG62209.
XX
PT   Novel 27 isolated human secreted proteins and polynucleotides encoding
PT   them useful for treating, diagnosing, preventing Alzheimer's disease,
PT   Parkinson's disease, AIDS, rheumatoid arthritis, asthma  -
XX
PS   Claim 4; Page 478-479; 594pp; English.
XX
CC   AAH19170-AAH19250 represent cDNAs corresponding to 27 human secreted
CC   protein genes, and AAG62156-AAG62235 represent the proteins they encode.
CC   AAG62236-AAG62293 represent human secreted protein fragments. The genes
CC   and their corresponding secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   52 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinsons's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 2319 BP; 431 A; 696 C; 736 G; 456 T; 0 other;


  Query Match             56.0%;  Score 14;  DB 22;  Length 2319;
  Best Local Similarity   100.0%;  Pred. No. 65;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db   1993 TCTGCTGTTTGAAG 1980


RESULT  39
AAD07655/c
ID   AAD07655 standard; cDNA; 2339 BP.
XX
AC   AAD07655;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 1 cDNA clone HDPVW11, SEQ ID NO:11.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW   psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW   inflammation; neurological disorder; Alzheimer's disease; food additive;
KW   angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW   pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW   cell culture; chemotaxis; vulnerary; binding partner identification;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             67..1434
FT                   /*tag=  a
FT                   /product= "Human secreted protein precursor"
FT   sig_peptide     67..150
FT                   /*tag=  b
FT   mat_peptide     151..1431
FT                   /*tag=  c
FT                   /product= "Mature human secreted protein"
XX
PN   WO200134644-A1.
XX
PD   17-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US30679.
XX
PR   12-NOV-1999;   99US-0164834.
PR   04-AUG-2000; 2000US-0224007.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Olsen HS,  Duan RD,  Ebner R;
XX
DR   WPI; 2001-329070/34.
DR   P-PSDB; AAE03202.
XX
PT   Isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 1; Page 394-395; 499pp; English.
XX
CC   AAD07655-AAD07695 represent cDNAs corresponding to 15 human secreted
CC   protein genes, and AAE03202-AAE03242 represent the proteins they encode.
CC   AAE03243-AAE03280 represent human secreted protein fragments or variants.
```

CC The secreted proteins and their genes are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 15 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein-encoding cDNA of
CC the invention.
XX
SQ Sequence 2339 BP; 477 A; 691 C; 720 G; 451 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2339;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tctgctgtttgaag 17
||||||||||||||
Db 1959 TCTGCTGTTTGAAG 1946

RESULT 40
AAF89054/c
ID AAF89054 standard; DNA; 2402 BP.
XX
AC AAF89054;
XX
DT 06-JUL-2001 (first entry)
XX
DE Human FATP3 coding sequence SEQ ID NO: 116.
XX
KW Fatty acid transport protein; FATP; human; mouse; rat; rice blast fungus;
KW yeast; fat absorption; obesity; diabetes; heart disease; hyperlipidaemia;
KW weight control; tuberculosis; TB; anti-fungal; ds.
XX
OS Homo sapiens.
XX
PN WO200121795-A2.
XX
PD 29-MAR-2001.
XX
PF 21-SEP-2000; 2000WO-US25891.
XX
PR 23-SEP-1999; 99US-0405504.
PR 23-SEP-1999; 99US-0405505.
PR 16-DEC-1999; 99US-0465280.
PR 17-FEB-2000; 2000US-0506252.
PR 06-JUL-2000; 2000US-0611197.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Stahl A, Hirsch DJ, Lodish HF, Gimeno RE, Tartaglia LA;
XX
DR WPI; 2001-354783/37.
DR P-PSDB; AAB83279.
XX
PT New fatty acid transport proteins (FATPs) useful for the manufacture of
PT medicament for treating obesity, diabetes and heart disease -
XX
PS Claim 1; Fig 111; 287pp; English.
XX
CC The present invention provides the protein and coding sequences of fatty
CC acid transport proteins (FATPs) from a number of species, including
CC FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6 from the human, FATP1-FATP5
CC from the mouse, FATPa and b from C. elegans, and FATP from Aspergillus
CC nidulans, Drosophila, zebrafish, Magnaporthe grisea, Mycobacterium
CC tuberculosis and Cochliobolus heterostrophus. The FATP from M.
CC tuberculosis can be used to identify inhibitors which can then be used to
CC treat TB. That from M. grisea (also known as rice blast fungus) can be
CC used to develop anti-fungal agents capable of preventing infection of
CC rice. Those from the human can be used to develop treatments for
CC diabetes, heart disease, obesity, hyperlipidaemia and weight control. The
CC present sequence is one of the sequences described in the exemplification
CC of the invention.
XX
SQ Sequence 2402 BP; 418 A; 728 C; 780 G; 476 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2402;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tctgctgtttgaag 17
||||||||||||||
Db 2103 TCTGCTGTTTGAAG 2090

RESULT 41
AAF27417/c
ID AAF27417 standard; cDNA; 2405 BP.
XX
AC AAF27417;
XX
DT 24-APR-2001 (first entry)
XX
DE Human fatty acid transporter PSEC67 cDNA.
XX
KW Human; fatty acid transporter; PSEC67; long-chain fatty acid uptake;
KW oleic acid; drug screening; gene therapy; metabolic disorder;
KW cardiomyopathy; skeletal muscle disorders; renal failure; ss.
XX
OS Homo sapiens.
XX
PN WO200104301-A1.
XX
PD 18-JAN-2001.
XX
PF 07-JUL-2000; 2000WO-JP04549.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 18-OCT-1999; 99US-0159586.
PR 25-APR-2000; 2000JP-0128993.
XX
PA (HELI-) HELIX RES INST.
XX
PI Morikawa N, Masuho Y, Ota T, Isogai T, Nishikawa T, Kawai Y;
XX
DR WPI; 2001-138349/14.
DR P-PSDB; AAB60388.
XX
PT Fatty acid transporter protein and encoded gene PSEC67 cloned from
PT human cDNA library, with activity of oleic acid incorporation, useful
PT as target molecule of preventives or remedies of fatty-acid metabolic
PT disorders -
XX
PS Claim 1; Page 44-48; 58pp; Japanese.
XX

CC The invention relates to a novel human fatty acid transporter,
CC PSEC67 (AAB60388), and to cDNA encoding it (AAF27417). PSEC67 is
CC responsible for the uptake of oleic acid into cells. The invention
CC also relates to vectors and host cells comprising a PSEC67
CC nucleic acid; the recombinant production of PSEC67; an antibody
CC against PSEC67; methods of screening for compounds which can regulate
CC the uptake of long-chain fatty acids into cells; and the compounds thus
CC identified. The PSEC67 protein and the gene encoding it are useful as
CC targets for the treatment or prevention of diseases associated with
CC defective metabolism of long-chain fatty acids. Such diseases include
CC cardiomyopathy, skeletal muscle disorders, or renal failure. PSEC67
CC nucleic acids may also be used in gene therapy for such disorders.
CC The present sequence represents cDNA encoding PSEC67.
XX
SQ Sequence 2405 BP; 410 A; 735 C; 781 G; 479 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2405;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tctgctgtttgaag 17
||||||||||||||
Db 2111 TCTGCTGTTTGAAG 2098

RESULT 42
AAH19186/c
ID AAH19186 standard; cDNA; 2447 BP.
XX
AC AAH19186;
XX
DT 17-JUL-2001 (first entry)
XX
DE Human secreted protein-encoding gene 17 cDNA clone HSLCA15, SEQ ID NO:27.
XX
KW Human; secreted protein; proliferative disorder; cancer;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder; tumour;
KW endocrine disorder; infection; wound healing; vulnerary;
KW cell culture; chemotaxis; food additive;
KW binding partner identification; ss.
XX
OS Homo sapiens.
XX
PN WO200132910-A2.
XX
PD 10-MAY-2001.
XX
PF 25-OCT-2000; 2000WO-US29362.
XX
PR 29-OCT-1999; 99US-0162240.
PR 30-JUN-2000; 2000US-0215131.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Birse CE, Ni J, Soppet DR;
XX
DR WPI; 2001-335835/35.
DR N-PSDB; AAG62172.
XX
PT Novel 27 isolated human secreted proteins and polynucleotides encoding
PT them useful for treating, diagnosing, preventing Alzheimer's disease,
PT Parkinson's disease, AIDS, rheumatoid arthritis, asthma -
XX
PS Claim 4; Page 456-457; 594pp; English.
XX
CC AAH19170-AAH19250 represent cDNAs corresponding to 27 human secreted
CC protein genes, and AAG62156-AAG62235 represent the proteins they encode.
CC AAG62236-AAG62293 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinsons's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 2447 BP; 444 A; 735 C; 781 G; 487 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 2447;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tctgctgtttgaag 17
||||||||||||||
Db 2111 TCTGCTGTTTGAAG 2098

RESULT 43
AAZ33977/c
ID AAZ33977 standard; cDNA; 2574 BP.
XX
AC AAZ33977;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO703 nucleotide sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein; ss.
XX
OS Homo sapiens.
XX
PN WO9946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.

PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
PA (GETH) GENENTECH INC.
XX
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
DR WPI; 1999-551358/46.
DR P-PSDB; AAY41699.
XX
PT New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
PS Claim 2; Fig 38; 530pp; English.
XX
CC The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AAZ33891 to
CC AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
CC invention.
XX
SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2574;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tctgctgtttgaag 17
||||||||||||||
Db 2167 TCTGCTGTTTGAAG 2154

RESULT 44
AAC78481/c
ID AAC78481 standard; cDNA; 2574 BP.
XX
AC AAC78481;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human PRO703 (UNQ367) nucleotide sequence SEQ ID NO:101.
XX
KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW expressed sequence tag; detection; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200053756-A2.
XX
PD 14-SEP-2000.
XX
PF 18-FEB-2000; 2000WO-US04341.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 14-MAY-1999; 99US-0134287.
PR 23-JUN-1999; 99US-0141037.
PR 26-JUL-1999; 99US-0145698.
PR 29-OCT-1999; 99US-0162506.

PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
DR WPI; 2000-611443/58.
DR P-PSDB; AAB44255.
XX
PT Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
PT cellular activities -
XX
PS Claim 2; Fig 38; 636pp; English.
XX
CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC sequence tag) sequences which encode secreted or transmembrane PRO
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC activity. The polynucleotides and polypeptides can be used for detecting
CC the presence of PRO polypeptides in samples, for linking bioactive
CC molecules to cells and for modulating biological activities of cells,
CC using the polypeptides for specific targeting. The polypeptide targeting
CC can be used to kill the target cells, e.g. for the treatment of cancers.
CC The polypeptide pairs provide specific targeting of bioactive molecules
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC the isolation of the PRO polynucleotide sequences.
XX
SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2574;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tctgctgtttgaag 17
||||||||||||||
Db 2167 TCTGCTGTTTGAAG 2154

RESULT 45
AAC58239/c
ID AAC58239 standard; cDNA; 2574 BP.
XX
AC AAC58239;
XX
DT 25-JAN-2001 (first entry)
XX
DE Human PRO703 nucleotide sequence SEQ ID NO:28.
XX
KW Human; tumour; diagnosis; neoplastic disease; identification; cancer;
KW tumourigenesis; detection; neoplastic cell growth; proliferation;
KW cytostatic; antiinflammatory; immunomodulatory; inflammatory disorder;
KW immunological disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200053754-A1.
XX
PD 14-SEP-2000.
XX
PF 06-JAN-2000; 2000WO-US00277.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 05-OCT-1999; 99WO-US23089.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Desauvage FJ, Goddard A, Gurney AL, Klein RD, Roy MA;
PI Wood WI;
XX
DR WPI; 2000-572269/53.
DR P-PSDB; AAB24054.
XX
PT New isolated antibody for use in compositions and methods for the
PT diagnosis and treatment of neoplastic cell growth and proliferation in
PT mammals, including humans, and in monitoring tumor treatment -
XX
PS Claim 50; Fig 28; 195pp; English.
XX
CC The present invention describes an isolated antibody (Ab) that binds to
CC one of the human proteins (P) designated PRO213, PRO1330, PRO1449,
CC PRO237, PRO324, PRO351, PRO362, PRO615, PRO531, PRO538, PRO3664, PRO618,
CC PRO772, PRO703, PRO792 or PRO474. The Ab can be used in compositions
CC and methods for the diagnosis and treatment of neoplastic cell growth
CC and proliferation in mammals, including humans. Genes and polypeptides
CC encoded by them, that are amplified in the genome of a tumour cell, can
CC be identified and are useful targets for the treatment and prevention of
CC certain cancers and may be used to monitor tumour treatment. Compounds
CC that inhibit the expression or activity of the identified polypeptides
CC can be identified and used as antagonists. Benign or malignant tumours,
CC inflammatory disorders and immunological disorders can be treated.
CC AAC58123 to AAC58224 represent hybridisation probes and PCR primers used
CC in the isolation of the human PRO sequences. AAC58225 to AAC58241 and
CC AAB24041 to AAB24056 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX
SQ Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2574;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tctgctgtttgaag 17
||||||||||||||
Db 2167 TCTGCTGTTTGAAG 2154

Search completed: December 26, 2001, 14:16:40
Job time: 6992 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 12:52:35 ; Search time 143.1 Seconds
(without alignments)
39.566 Million cell updates/sec

Title: US-09-396-196F-4
Perfect score: 25
Sequence: 1 ggatctgctgtttgaagcgcagcag 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 9

Total number of hits satisfying chosen parameters: 12481

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		25	100.0	1041	2	US-08-401-068-7	Sequence 7, Appli
2		25	100.0	1041	2	US-08-846-338-7	Sequence 7, Appli
3		25	100.0	5872	3	US-08-411-768B-1	Sequence 1, Appli
4		25	100.0	5872	3	US-08-411-768B-6	Sequence 6, Appli
5	c	14	56.0	753	4	US-09-232-191-14	Sequence 14, Appl
6	c	14	56.0	753	4	US-09-232-200-14	Sequence 14, Appl
7	c	14	56.0	753	4	US-09-232-197-14	Sequence 14, Appl
8	c	14	56.0	1173	4	US-09-232-200-50	Sequence 50, Appl
9	c	14	56.0	1173	4	US-09-232-197-50	Sequence 50, Appl
10	c	14	56.0	1998	4	US-09-232-200-68	Sequence 68, Appl
11	c	14	56.0	1998	4	US-09-232-197-68	Sequence 68, Appl
12	c	14	56.0	2087	4	US-09-232-191-6	Sequence 6, Appli
13	c	14	56.0	2087	4	US-09-232-200-6	Sequence 6, Appli
14	c	14	56.0	2087	4	US-09-232-197-6	Sequence 6, Appli
15	c	14	56.0	2166	4	US-09-232-200-101	Sequence 101, App
16	c	14	56.0	2166	4	US-09-232-197-101	Sequence 101, App
17		14	56.0	12284	2	US-08-876-991-1	Sequence 1, Appli
18		14	56.0	12284	2	US-09-059-853-1	Sequence 1, Appli
19	c	13	52.0	17	1	US-08-233-608-8	Sequence 8, Appli
20	c	13	52.0	17	1	US-08-887-480-8	Sequence 8, Appli
21	c	13	52.0	17	2	US-08-722-187-8	Sequence 8, Appli
22	c	13	52.0	17	5	PCT-US95-04712-8	Sequence 8, Appli
23	c	13	52.0	23	1	US-08-233-608-9	Sequence 9, Appli
24	c	13	52.0	23	1	US-08-887-480-9	Sequence 9, Appli
25	c	13	52.0	23	2	US-08-722-187-9	Sequence 9, Appli
26	c	13	52.0	23	5	PCT-US95-04712-9	Sequence 9, Appli
27	c	13	52.0	27	1	US-08-150-331-31	Sequence 31, Appl
28	c	13	52.0	384	2	US-08-820-170A-8	Sequence 8, Appli
29	c	13	52.0	384	3	US-09-055-699-8	Sequence 8, Appli
30	c	13	52.0	384	4	US-09-273-565-8	Sequence 8, Appli
31		13	52.0	583	1	US-08-233-608-2	Sequence 2, Appli
32		13	52.0	583	1	US-08-887-480-2	Sequence 2, Appli
33		13	52.0	583	2	US-08-722-187-2	Sequence 2, Appli
34		13	52.0	583	5	PCT-US95-04712-2	Sequence 2, Appli
35		13	52.0	749	4	US-08-998-416-727	Sequence 727, App
36		13	52.0	1161	1	US-08-153-848-31	Sequence 31, Appl
37		13	52.0	1161	5	PCT-US93-11153-31	Sequence 31, Appl
38	c	13	52.0	1242	1	US-08-252-966B-13	Sequence 13, Appl
39	c	13	52.0	1425	1	US-08-353-550-7	Sequence 7, Appli
40	c	13	52.0	1425	2	US-08-551-687-7	Sequence 7, Appli
41	c	13	52.0	1493	2	US-08-820-170A-9	Sequence 9, Appli
42	c	13	52.0	1493	3	US-09-055-699-9	Sequence 9, Appli
43	c	13	52.0	1493	4	US-09-273-565-9	Sequence 9, Appli
44		13	52.0	1712	1	US-08-171-299B-1	Sequence 1, Appli
45	c	13	52.0	1743	3	US-09-221-235-9	Sequence 9, Appli
46	c	13	52.0	1743	3	US-09-221-928-9	Sequence 9, Appli
47	c	13	52.0	1743	3	US-09-221-527-9	Sequence 9, Appli
48	c	13	52.0	1743	3	US-09-221-236-9	Sequence 9, Appli
49	c	13	52.0	1743	3	US-09-221-416-9	Sequence 9, Appli
50	c	13	52.0	1743	4	US-09-221-245-9	Sequence 9, Appli
51	c	13	52.0	1743	4	US-09-163-115-9	Sequence 9, Appli
52	c	13	52.0	1743	4	US-09-221-528-9	Sequence 9, Appli
53	c	13	52.0	1743	4	US-09-593-553-9	Sequence 9, Appli
54	c	13	52.0	1743	4	US-09-221-237-9	Sequence 9, Appli
55		13	52.0	1835	1	US-08-361-611-3	Sequence 3, Appli
56		13	52.0	1835	1	US-08-565-655-3	Sequence 3, Appli
57		13	52.0	1835	2	US-08-946-967-3	Sequence 3, Appli
58		13	52.0	2143	4	US-09-064-693A-18	Sequence 18, Appl
59		13	52.0	2254	1	US-08-153-848-27	Sequence 27, Appl
60		13	52.0	2254	3	US-09-299-843A-27	Sequence 27, Appl
61		13	52.0	2254	5	PCT-US93-11153-27	Sequence 27, Appl
62	c	13	52.0	2310	1	US-08-416-581B-2	Sequence 2, Appli
63	c	13	52.0	2454	3	US-09-221-235-7	Sequence 7, Appli
64	c	13	52.0	2454	3	US-09-221-928-7	Sequence 7, Appli
65	c	13	52.0	2454	3	US-09-221-527-7	Sequence 7, Appli
66	c	13	52.0	2454	3	US-09-221-236-7	Sequence 7, Appli
67	c	13	52.0	2454	3	US-09-221-416-7	Sequence 7, Appli
68	c	13	52.0	2454	4	US-09-221-245-7	Sequence 7, Appli
69	c	13	52.0	2454	4	US-09-163-115-7	Sequence 7, Appli
70	c	13	52.0	2454	4	US-09-221-528-7	Sequence 7, Appli
71	c	13	52.0	2454	4	US-09-593-553-7	Sequence 7, Appli
72	c	13	52.0	2454	4	US-09-221-237-7	Sequence 7, Appli
73	c	13	52.0	2674	2	US-08-926-724-2	Sequence 2, Appli
74	c	13	52.0	2787	1	US-08-416-581B-3	Sequence 3, Appli
75	c	13	52.0	2787	1	US-08-416-581B-4	Sequence 4, Appli
76	c	13	52.0	2787	3	US-09-288-461-1	Sequence 1, Appli
77	c	13	52.0	2787	4	US-09-087-465-5	Sequence 5, Appli
78	c	13	52.0	3031	4	US-08-664-962B-3	Sequence 3, Appli
79	c	13	52.0	3031	4	US-09-311-743-3	Sequence 3, Appli
80		13	52.0	3119	3	US-09-299-843A-31	Sequence 31, Appl
81		13	52.0	3435	1	US-08-366-577-1	Sequence 1, Appli
82		13	52.0	3435	5	PCT-US96-00005-1	Sequence 1, Appli
83		13	52.0	4856	4	US-09-045-360-1	Sequence 1, Appli
84	c	13	52.0	4954	1	US-08-920-812-18	Sequence 18, Appl
85	c	13	52.0	4954	1	US-08-920-827-18	Sequence 18, Appl
86	c	13	52.0	4954	1	US-08-921-177-18	Sequence 18, Appl
87	c	13	52.0	4954	1	US-08-362-577C-18	Sequence 18, Appl
88	c	13	52.0	4954	2	US-08-920-828-18	Sequence 18, Appl
89	c	13	52.0	5312	4	US-08-485-355B-39	Sequence 39, Appl
90	c	13	52.0	5312	4	US-08-485-355B-41	Sequence 41, Appl
91	c	13	52.0	5312	4	US-08-485-355B-43	Sequence 43, Appl
92	c	13	52.0	5368	4	US-08-485-355B-45	Sequence 45, Appl
93		13	52.0	6641	4	US-09-064-693A-25	Sequence 25, Appl
94	c	13	52.0	24979	2	US-08-147-777-3	Sequence 3, Appli
95	c	13	52.0	24979	3	US-08-452-872-3	Sequence 3, Appli
96	c	13	52.0	24979	5	PCT-US93-03985-3	Sequence 3, Appli
97	c	13	52.0	46899	1	US-08-471-119A-1	Sequence 1, Appli
98	c	12	48.0	25	2	US-08-257-781-19	Sequence 19, Appl
99	c	12	48.0	25	5	PCT-US95-06857-19	Sequence 19, Appl
100		12	48.0	30	2	US-08-766-439-87	Sequence 87, Appl

ALIGNMENTS

```
RESULT   1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Enhanced Biotin Biosynthesis in Plant Tissue
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/401,068
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/351,970
;      FILING DATE:  08-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-401-068-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db      63 GGATCTGCTGTTTGAAGCGCAGCAG 87


RESULT   2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
;  GENERAL INFORMATION:
;    APPLICANT:  Patton, David
;    TITLE OF INVENTION:  Transgenic Plants Having Increased Biotin Content
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5869719artis Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/846,338
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1038
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /product= "biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
US-08-846-338-7


  Query Match             100.0%;  Score 25;  DB 2;  Length 1041;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaagcgcagcag 25
           |||||||||||||||||||||||||
Db      63 GGATCTGCTGTTTGAAGCGCAGCAG 87


RESULT   3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A-15/9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1157
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 117
;      OTHER INFORMATION:  /product= "Biotin synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioB"
;      OTHER INFORMATION:  /number= 1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2295..3050
;      OTHER INFORMATION:  /codon_start= 2295
;      OTHER INFORMATION:  /function= "involved in pimeloyl-CoA synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /gene= "bioC"
;      OTHER INFORMATION:  /number= 3
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3750..5039
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3750
;      OTHER INFORMATION:  /EC_number= 2.6.1.62
;      OTHER INFORMATION:  /product= "DAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioA"
;      OTHER INFORMATION:  /number= 5
;      OTHER INFORMATION:  /standard_name=
;      OTHER INFORMATION:  "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
;      OTHER INFORMATION:  aminotransf."
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  5098..5574
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 5098
;      OTHER INFORMATION:  /function= "unknown, involved in biotin synthesis"
;      OTHER INFORMATION:  /product= "protein"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "ORFI"
;      OTHER INFORMATION:  /number= 6
;    FEATURE:
;      NAME/KEY:  -10_signal
;      LOCATION:  45..49
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  -35_signal
;      LOCATION:  23..28
;      OTHER INFORMATION:  /standard_name= "promoter ptac"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  105..119
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "bioB RBS no.9"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  2284..2297
;      OTHER INFORMATION:  /standard_name= "bioC RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3742..3752
;      OTHER INFORMATION:  /standard_name= "bioA RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  5088..5100
;      OTHER INFORMATION:  /standard_name= "ORFI RBS"
;    FEATURE:
;      NAME/KEY:  terminator
;      LOCATION:  5583..5644
;      OTHER INFORMATION:  /standard_name= "rho-independent
;      OTHER INFORMATION:  transcriptional terminator"
;    FEATURE:
;      NAME/KEY:  stem_loop
;      LOCATION:  5583..5605
;    FEATURE:
;      NAME/KEY:  promoter
;      LOCATION:  1..96
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "promoter ptac"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-1


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagcgcagcag 25
          |||||||||||||||||||||||||
Db    179 GGATCTGCTGTTTGAAGCGCAGCAG 203


RESULT   4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
;  GENERAL INFORMATION:
;    APPLICANT:  Olwen Birch
;    APPLICANT:  Johann Brass
;    APPLICANT:  Martin Fuhrmann
;    APPLICANT:  Nicholas Shaw
;    TITLE OF INVENTION:  Biotechnological Method
;    TITLE OF INVENTION:  of Producing Biotin
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect
;      SOFTWARE:  Version 5.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/411,768B
;      FILING DATE:  31-March-95
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 3124/92
;      FILING DATE:  02-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 2134/93
;      FILING DATE:  15-JUL-1993
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5872 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Escherichia coli
;      STRAIN:  DSM498
;    IMMEDIATE SOURCE:
;      CLONE:  pBO30A15-9
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1154..2308
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 1154
;      OTHER INFORMATION:  /EC_number= 2.3.1.47
;      OTHER INFORMATION:  /product= "KAPA synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioF"
;      OTHER INFORMATION:  /number= 2
;      OTHER INFORMATION:  /standard_name= "8-Amino-7-oxononanoate synthase"
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3043..3753
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /codon_start= 3043
;      OTHER INFORMATION:  /EC_number= 6.3.3.3
;      OTHER INFORMATION:  /product= "DTB synthase"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /gene= "bioD"
;      OTHER INFORMATION:  /number= 4
;      OTHER INFORMATION:  /standard_name= "Dethiobiotin synthase"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  1141..1156
;      OTHER INFORMATION:  /standard_name= "bioF RBS"
;    FEATURE:
;      NAME/KEY:  RBS
;      LOCATION:  3030..3045
;      OTHER INFORMATION:  /standard_name= "bioD RBS"
;    PUBLICATION INFORMATION:
;      DOCUMENT NUMBER:  WO 87/01391 B1
;      FILING DATE:  26-AUG-1986
;      PUBLICATION DATE:  07-APR-1993
US-08-411-768B-6


  Query Match             100.0%;  Score 25;  DB 3;  Length 5872;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttgaagcgcagcag 25
          |||||||||||||||||||||||||
Db    179 GGATCTGCTGTTTGAAGCGCAGCAG 203


RESULT   5
US-09-232-191-14/c
; Sequence 14, Application US/09232191
; Patent No. 6284487
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  TITLE OF INVENTION: Fatty Acid Transport Proteins
;  FILE REFERENCE: WHI97-21p3ME
;  CURRENT APPLICATION NUMBER: US/09/232,191
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-232-191-14


  Query Match             56.0%;  Score 14;  DB 4;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db    436 TCTGCTGTTTGAAG 423


RESULT   6
US-09-232-200-14/c
; Sequence 14, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-232-200-14


  Query Match             56.0%;  Score 14;  DB 4;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db    436 TCTGCTGTTTGAAG 423


RESULT   7
US-09-232-197-14/c
; Sequence 14, Application US/09232197A
```

```
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 14
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-232-197-14


  Query Match             56.0%;  Score 14;  DB 4;  Length 753;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db     436 TCTGCTGTTTGAAG 423


RESULT   8
US-09-232-200-50/c
; Sequence 50, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 50
;   LENGTH: 1173
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-232-200-50


  Query Match             56.0%;  Score 14;  DB 4;  Length 1173;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db     853 TCTGCTGTTTGAAG 840


RESULT   9
US-09-232-197-50/c
; Sequence 50, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 50
;   LENGTH: 1173
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-232-197-50


  Query Match             56.0%;  Score 14;  DB 4;  Length 1173;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db     853 TCTGCTGTTTGAAG 840


RESULT  10
US-09-232-200-68/c
; Sequence 68, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 68
;   LENGTH: 1998
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-200-68


  Query Match             56.0%;  Score 14;  DB 4;  Length 1998;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
```

```
Db    1690 TCTGCTGTTTGAAG 1677


RESULT  11
US-09-232-197-68/c
; Sequence 68, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 68
;   LENGTH: 1998
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-197-68


  Query Match             56.0%;  Score 14;  DB 4;  Length 1998;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db    1690 TCTGCTGTTTGAAG 1677


RESULT  12
US-09-232-191-6/c
; Sequence 6, Application US/09232191
; Patent No. 6284487
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  TITLE OF INVENTION: Fatty Acid Transport Proteins
;  FILE REFERENCE: WHI97-21p3ME
;  CURRENT APPLICATION NUMBER: US/09/232,191
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 2087
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-191-6


  Query Match             56.0%;  Score 14;  DB 4;  Length 2087;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db    1779 TCTGCTGTTTGAAG 1766


RESULT  13
US-09-232-200-6/c
; Sequence 6, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 2087
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-200-6


  Query Match             56.0%;  Score 14;  DB 4;  Length 2087;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaag 17
          ||||||||||||||
Db    1779 TCTGCTGTTTGAAG 1766


RESULT  14
US-09-232-197-6/c
; Sequence 6, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 2087
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-232-197-6


  Query Match             56.0%;  Score 14;  DB 4;  Length 2087;
  Best Local Similarity   100.0%;  Pred. No. 17;
```

```
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db    1779 TCTGCTGTTTGAAG 1766


RESULT  15
US-09-232-200-101/c
; Sequence 101, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MB
;  CURRENT APPLICATION NUMBER: US/09/232,200A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 101
;   LENGTH: 2166
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (19)...(2124)
US-09-232-200-101


  Query Match             56.0%;  Score 14;  DB 4;  Length 2166;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db    1987 TCTGCTGTTTGAAG 1974


RESULT  16
US-09-232-197-101/c
; Sequence 101, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
;  APPLICANT: Stahl, Andreas
;  APPLICANT: Hirsch, David J.
;  APPLICANT: Lodish, Harvey F.
;  APPLICANT: Gimeno, Ruth E.
;  APPLICANT: Tartaglia, Louis A.
;  TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
;  FILE REFERENCE: WHI97-21p3MA
;  CURRENT APPLICATION NUMBER: US/09/232,197A
;  CURRENT FILING DATE: 1999-01-14
;  EARLIER APPLICATION NUMBER: 60/071,374
;  EARLIER FILING DATE: 1998-01-15
;  EARLIER APPLICATION NUMBER: 60/093,491
;  EARLIER FILING DATE: 1998-07-20
;  EARLIER APPLICATION NUMBER: 60/110,941
;  EARLIER FILING DATE: 1998-12-04
;  NUMBER OF SEQ ID NOS: 105
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 101
;   LENGTH: 2166
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (19)...(2124)
US-09-232-197-101


  Query Match             56.0%;  Score 14;  DB 4;  Length 2166;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db    1987 TCTGCTGTTTGAAG 1974


RESULT  17
US-08-876-991-1
; Sequence 1, Application US/08876991
; Patent No. 5925360
;  GENERAL INFORMATION:
;    APPLICANT:  Gregor Meyers, Tillmann R menapf,
;    APPLICANT:  Heinz-J rgen Thiel
;    TITLE OF INVENTION:  Hog cholera virus vaccine and diagnostic
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Organon Teknika Corporation
;      ADDRESSEE:  Biotechnology Research Institute
;      STREET:  1330-A Piccard Drive
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  U.S.A.
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,991
;      FILING DATE:  16-JUN-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,577
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/650,584
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/469,702
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/123,596
;      FILING DATE:
;      APPLICATION NUMBER:  07/797,554
;      FILING DATE:  22-NOV-1991
;      APPLICATION NUMBER:  US 07/494,991
;      FILING DATE:  16-MAR-1990
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William M. Blackstone
;      REGISTRATION NUMBER:  29,772
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 258-5200
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12284 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Hog cholera virus
;      STRAIN:  Alfort
;      CELL LINE:  PK 15  and 38A1D
```

```
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  364..12060
;       OTHER INFORMATION:  /label= 435_kDA_protein
;     FEATURE:
;       NAME/KEY:  primer_bind
;       LOCATION:  complement (2587..2619)
;       OTHER INFORMATION:  /label= primer_1
;     FEATURE:
;       NAME/KEY:  primer_bind
;       LOCATION:  complement (2842..2880)
;       OTHER INFORMATION:  /label= primer_2
;     FEATURE:
;       NAME/KEY:  variation
;       LOCATION:  replace(127, "c")
;     FEATURE:
;       NAME/KEY:  variation
;       LOCATION:  replace(1522, "g")
;     FEATURE:
;       NAME/KEY:  variation
;       LOCATION:  replace(10989, "t")
US-08-876-991-1


  Query Match             56.0%;  Score 14;  DB 2;  Length 12284;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db    2253 TCTGCTGTTTGAAG 2266


RESULT  18
US-09-059-853-1
; Sequence 1, Application US/09059853
; Patent No. 5935582
;  GENERAL INFORMATION:
;    APPLICANT:  Gregor Meyers, Tillmann R menapf,
;    APPLICANT:  Heinz-J rgen Thiel
;    TITLE OF INVENTION:  Hog cholera virus vaccine and diagnostic
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Organon Teknika Corporation
;      ADDRESSEE:  Biotechnology Research Institute
;      STREET:  1330-A Piccard Drive
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  U.S.A.
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/059,853
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/797,554
;      FILING DATE:  22-NOV-1991
;      APPLICATION NUMBER:  US 07/494,991
;      FILING DATE:  16-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William M. Blackstone
;      REGISTRATION NUMBER:  29,772
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 258-5200
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12284 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Hog cholera virus
;      STRAIN:  Alfort
;      CELL LINE:  PK 15  and 38A1D
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  364..12060
;      OTHER INFORMATION:  /label= 435_kDA_protein
;    FEATURE:
;      NAME/KEY:  primer_bind
;      LOCATION:  complement (2587..2619)
;      OTHER INFORMATION:  /label= primer_1
;    FEATURE:
;      NAME/KEY:  primer_bind
;      LOCATION:  complement (2842..2880)
;      OTHER INFORMATION:  /label= primer_2
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(127, "c")
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(1522, "g")
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(10989, "t")
US-09-059-853-1


  Query Match             56.0%;  Score 14;  DB 2;  Length 12284;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaag 17
           ||||||||||||||
Db    2253 TCTGCTGTTTGAAG 2266


RESULT  19
US-08-233-608-8/c
; Sequence 8, Application US/08233608
; Patent No. 5585238
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,608
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
```

```
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  17 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Other nucleic acid
;       DESCRIPTION:  Oligonucleotide primer JB434
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-233-608-8


  Query Match             52.0%;  Score 13;  DB 1;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db      15 TTGAAGCGCAGCA 3


RESULT  20
US-08-887-480-8/c
; Sequence 8, Application US/08887480
; Patent No. 5814453
;  GENERAL INFORMATION:
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  96
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5814453artis Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/887,480
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,187
;      FILING DATE:  15-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1739/PCT/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other nucleic acid
;      DESCRIPTION:  Oligonucleotide primer JB434
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-887-480-8


  Query Match             52.0%;  Score 13;  DB 1;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db      15 TTGAAGCGCAGCA 3


RESULT  21
US-08-722-187-8/c
; Sequence 8, Application US/08722187
; Patent No. 5955274
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,187
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/233,608
;      FILING DATE:  04-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Walsh, Andrea C.
;      REGISTRATION NUMBER:  34,988
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8666
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other nucleic acid
;      DESCRIPTION:  Oligonucleotide primer JB434
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-722-187-8


  Query Match             52.0%;  Score 13;  DB 2;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db      15 TTGAAGCGCAGCA 3


RESULT  22
PCT-US95-04712-8/c
; Sequence 8, Application PC/TUS9504712
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
```

```
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04712
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/233,608
;      FILING DATE:  04-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Walsh, Andrea C.
;      REGISTRATION NUMBER:  34,988
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8666
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other nucleic acid
;      DESCRIPTION:  Oligonucleotide primer JB434
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US95-04712-8


  Query Match             52.0%;  Score 13;  DB 5;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db      15 TTGAAGCGCAGCA 3


RESULT  23
US-08-233-608-9/c
; Sequence 9, Application US/08233608
; Patent No. 5585238
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,608
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other nucleic acid
;      DESCRIPTION:  Oligonucleotide primer JB525
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-233-608-9


  Query Match             52.0%;  Score 13;  DB 1;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db      21 TTGAAGCGCAGCA 9


RESULT  24
US-08-887-480-9/c
; Sequence 9, Application US/08887480
; Patent No. 5814453
;  GENERAL INFORMATION:
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  96
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5814453artis Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/887,480
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,187
;      FILING DATE:  15-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1739/PCT/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other nucleic acid
```

```
;      DESCRIPTION:  Oligonucleotide primer JB525
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-887-480-9


  Query Match             52.0%;  Score 13;  DB 1;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db      21 TTGAAGCGCAGCA 9


RESULT  25
US-08-722-187-9/c
; Sequence 9, Application US/08722187
; Patent No. 5955274
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,187
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/233,608
;      FILING DATE:  04-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Walsh, Andrea C.
;      REGISTRATION NUMBER:  34,988
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8666
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other nucleic acid
;      DESCRIPTION:  Oligonucleotide primer JB525
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-722-187-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db      21 TTGAAGCGCAGCA 9


RESULT  26
PCT-US95-04712-9/c
; Sequence 9, Application PC/TUS9504712
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04712
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/233,608
;      FILING DATE:  04-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Walsh, Andrea C.
;      REGISTRATION NUMBER:  34,988
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8666
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  23 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Other nucleic acid
;      DESCRIPTION:  Oligonucleotide primer JB525
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
PCT-US95-04712-9


  Query Match             52.0%;  Score 13;  DB 5;  Length 23;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db      21 TTGAAGCGCAGCA 9


RESULT  27
US-08-150-331-31/c
; Sequence 31, Application US/08150331
; Patent No. 5516512
;  GENERAL INFORMATION:
;    APPLICANT:  DORSSERS J., LAMBERTUS C.
;    APPLICANT:  VAN LEEN, ROBERT W.
;    TITLE OF INVENTION:  MUTANTS OF HUMAN INTERLEUKIN-3
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1018
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/150,331
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/651,437
;      FILING DATE:  05-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  GRACEY, NANCY J.
;      REGISTRATION NUMBER:  28,216
;      REFERENCE/DOCKET NUMBER:  24615-20010.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-150-331-31


  Query Match             52.0%;  Score 13;  DB 1;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 53;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db      20 TTGAAGCGCAGCA 8


RESULT  28
US-08-820-170A-8/c
; Sequence 8, Application US/08820170A
; Patent No. 5831058
;  GENERAL INFORMATION:
;    APPLICANT:  Tsutomu, FUJIWARA
;    APPLICANT:  Takeshi, WATANABE
;    APPLICANT:  Masato, HORIE
;    APPLICANT:  Toyomasa, KATAGIRI
;    TITLE OF INVENTION:  HUMAN GENE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/820,170A
;      FILING DATE:
;      CLASSIFICATION:  536
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  384 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA(genomic)
US-08-820-170A-8


  Query Match             52.0%;  Score 13;  DB 2;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgttt 13
           |||||||||||||
Db     349 GGATCTGCTGTTT 337


RESULT  29
US-09-055-699-8/c
; Sequence 8, Application US/09055699
; Patent No. 6005088
;  GENERAL INFORMATION:
;    APPLICANT:  Tsutomu, FUJIWARA
;    APPLICANT:  Takeshi, WATANABE
;    APPLICANT:  Masato, HORIE
;    APPLICANT:  Toyomasa, KATAGIRI
;    TITLE OF INVENTION:  HUMAN GENE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/055,699
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/820,170
;      FILING DATE:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  384 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA(genomic)
US-09-055-699-8


  Query Match             52.0%;  Score 13;  DB 3;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgttt 13
           |||||||||||||
Db     349 GGATCTGCTGTTT 337


RESULT  30
US-09-273-565-8/c
; Sequence 8, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
```

```
;  APPLICANT: FUJIWARA, TSUTOMU
;  APPLICANT: WATANABE, TAKESHI
;  APPLICANT: HORIE, MASATO
;  TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
;  TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
;  FILE REFERENCE: Q-53599
;  CURRENT APPLICATION NUMBER: US/09/273,565A
;  CURRENT FILING DATE: 1999-03-22
;  EARLIER APPLICATION NUMBER: 09/055,699
;  EARLIER FILING DATE: 1998-04-07
;  EARLIER APPLICATION NUMBER: 08/820,170
;  EARLIER FILING DATE: 1997-03-19
;  EARLIER APPLICATION NUMBER: JP 63410/1996
;  EARLIER FILING DATE: 1996-03-19
;  EARLIER APPLICATION NUMBER: JP 69163/1997
;  EARLIER FILING DATE: 1997-03-05
;  NUMBER OF SEQ ID NOS: 95
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 384
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-273-565-8


  Query Match             52.0%;  Score 13;  DB 4;  Length 384;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgttt 13
          |||||||||||||
Db    349 GGATCTGCTGTTT 337


RESULT  31
US-08-233-608-2
; Sequence 2, Application US/08233608
; Patent No. 5585238
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/233,608
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spruill, W. Murray
;      REGISTRATION NUMBER:  32,943
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8615
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  583 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Septoria nodorum
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..583
;      OTHER INFORMATION:  /note= "DNA sequence for the
;      OTHER INFORMATION:  Internal Transcribed Spacer of Septoria nodorum"
US-08-233-608-2


  Query Match             52.0%;  Score 13;  DB 1;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db    464 TTGAAGCGCAGCA 476


RESULT  32
US-08-887-480-2
; Sequence 2, Application US/08887480
; Patent No. 5814453
;  GENERAL INFORMATION:
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  96
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5814453artis Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/887,480
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/722,187
;      FILING DATE:  15-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1739/PCT/CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  583 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Septoria nodorum
;    FEATURE:
;      NAME/KEY:  misc_ feature
;      LOCATION:  1..30
;      OTHER INFORMATION:  /note= "3' end of small subunit
;      OTHER INFORMATION:  rRNA gene"
```

```
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  31..216
;      OTHER INFORMATION:  /note= "ITS 1"
;    FEATURE:
;      NAME/KEY:  misc_ feature
;      LOCATION:  217..372
;      OTHER INFORMATION:  /note= "5.8S rRNA gene"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  373..526
;      OTHER INFORMATION:  /note= "ITS 2"
;    FEATURE:
;      NAME/KEY:  misc_ feature
;      LOCATION:  527..583
;      OTHER INFORMATION:  /note= "5' end of large subunit
;      OTHER INFORMATION:  rRNA gene"
US-08-887-480-2


  Query Match             52.0%;  Score 13;  DB 1;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db     464 TTGAAGCGCAGCA 476


RESULT  33
US-08-722-187-2
; Sequence 2, Application US/08722187
; Patent No. 5955274
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/722,187
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/233,608
;      FILING DATE:  04-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Walsh, Andrea C.
;      REGISTRATION NUMBER:  34,988
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8666
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  583 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Septoria nodorum
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..583
;      OTHER INFORMATION:  /note= "DNA sequence for the
;      OTHER INFORMATION:  Internal Transcribed Spacer of Septoria nodorum"
US-08-722-187-2


  Query Match             52.0%;  Score 13;  DB 2;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db     464 TTGAAGCGCAGCA 476


RESULT  34
PCT-US95-04712-2
; Sequence 2, Application PC/TUS9504712
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M
;    APPLICANT:  Beck, James J
;    TITLE OF INVENTION:  Detection of Fungal Pathogens Using the
;    TITLE OF INVENTION:  Polymerase Chain Reaction
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/04712
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/233,608
;      FILING DATE:  04-APR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Walsh, Andrea C.
;      REGISTRATION NUMBER:  34,988
;      REFERENCE/DOCKET NUMBER:  CGC 1739
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8666
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  583 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Septoria nodorum
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..583
;      OTHER INFORMATION:  /note= "DNA sequence for the
;      OTHER INFORMATION:  Internal Transcribed Spacer of Septoria nodorum"
PCT-US95-04712-2
```

```
  Query Match             52.0%;  Score 13;  DB 5;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 ttgaagcgcagca 24
           |||||||||||||
Db     464 TTGAAGCGCAGCA 476


RESULT  35
US-08-998-416-727
; Sequence 727, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  727:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  749 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  PAG1475RP
US-08-998-416-727


  Query Match             52.0%;  Score 13;  DB 4;  Length 749;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 tgctgtttgaagc 18
           |||||||||||||
Db      96 TGCTGTTTGAAGC 108


RESULT  36
US-08-153-848-31
; Sequence 31, Application US/08153848
; Patent No. 5759804
;  GENERAL INFORMATION:
;    APPLICANT:  Godiska, Ronald
;    APPLICANT:  Gray, Patrick W.
;    APPLICANT:  Schweikart, Vicki L.
;    TITLE OF INVENTION:  No. 5759804el Seven Transmembrane Receptors
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/153,848
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/977,452
;      FILING DATE:  17-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5759804and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  7..80
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  94..1158
US-08-153-848-31


  Query Match             52.0%;  Score 13;  DB 1;  Length 1161;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 gatctgctgtttg 14
           |||||||||||||
Db     322 GATCTGCTGTTTG 334


RESULT  37
PCT-US93-11153-31
; Sequence 31, Application PC/TUS9311153
;  GENERAL INFORMATION:
;    APPLICANT:  Godiska, Ronald
;    APPLICANT:  Gray, Patrick W.
;    APPLICANT:  Schweikart, Vicki L.
;    TITLE OF INVENTION:  Novel Seven Transmembrane Receptors
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
```

```
;       ADDRESSEE:  Bicknell
;       STREET:  6300 Sears Tower, 233 South Wacker Drive
;       CITY:  Chicago
;       STATE:  Illinois
;       COUNTRY:  USA
;       ZIP:  60606
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/11153
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/977,452
;       FILING DATE:  17-NOV-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Noland, Greta E.
;       REGISTRATION NUMBER:  35,302
;       REFERENCE/DOCKET NUMBER:  31794
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (312) 474-6300
;       TELEFAX:  (312) 474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  31:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1161 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  7..80
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  94..1158
PCT-US93-11153-31


  Query Match             52.0%;  Score 13;  DB 5;  Length 1161;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 gatctgctgtttg 14
           |||||||||||||
Db     322 GATCTGCTGTTTG 334


RESULT  38
US-08-252-966B-13/c
; Sequence 13, Application US/08252966B
; Patent No. 5624818
;  GENERAL INFORMATION:
;    APPLICANT:  Eisenman, Robert N.
;    APPLICANT:  Hurlin, Peter J.
;    APPLICANT:  Ayer, Donald E.
;    TITLE OF INVENTION:  Regulatory Proteins that Dimerize with
;    TITLE OF INVENTION:  Mad or Max
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen, O'Connor, Johnson, and KindnessPLLC
;      STREET:  1420 Fifth Ave., Suite 2800
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/252,966B
;      FILING DATE:  01-JUN-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shelton, Dennis K.
;      REGISTRATION NUMBER:  26,997.
;      REFERENCE/DOCKET NUMBER:  FHCR17694
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 682-8100
;      TELEFAX:  (206) 224-0779
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1242 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;      DESCRIPTION:  clone 10; see Figure 24
;    HYPOTHETICAL:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Mus musculus
US-08-252-966B-13


  Query Match             52.0%;  Score 13;  DB 1;  Length 1242;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ctgctgtttgaag 17
           |||||||||||||
Db     525 CTGCTGTTTGAAG 513


RESULT  39
US-08-353-550-7/c
; Sequence 7, Application US/08353550
; Patent No. 5744313
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Cavanaugh, William M.
;    TITLE OF INVENTION:  No. 5744313el Protein Domain Which Binds
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Proteins
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/353,550
;      FILING DATE:  09-DEC-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  02307K-057300
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1425 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1422
US-08-353-550-7


  Query Match             52.0%;  Score 13;  DB 1;  Length 1425;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 tgtttgaagcgca 21
           |||||||||||||
Db     611 TGTTTGAAGCGCA 599


RESULT  40
US-08-551-687-7/c
; Sequence 7, Application US/08551687
; Patent No. 5925547
;  GENERAL INFORMATION:
;    APPLICANT:  Williams, Lewis T.
;    APPLICANT:  Cavanaugh, William M.
;    TITLE OF INVENTION:  No. 5925547el Protein Domain Which Binds
;    TITLE OF INVENTION:  Tyrosine Phosphorylated Proteins
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/551,687
;      FILING DATE:  01-NOV-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/353,550
;      FILING DATE:  09-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Murphy, Matthew B.
;      REGISTRATION NUMBER:  P39,787
;      REFERENCE/DOCKET NUMBER:  2307K-5731
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1425 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1422
US-08-551-687-7


  Query Match             52.0%;  Score 13;  DB 2;  Length 1425;
  Best Local Similarity   100.0%;  Pred. No. 61;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 tgtttgaagcgca 21
           |||||||||||||
Db     611 TGTTTGAAGCGCA 599


RESULT  41
US-08-820-170A-9/c
; Sequence 9, Application US/08820170A
; Patent No. 5831058
;  GENERAL INFORMATION:
;    APPLICANT:  Tsutomu, FUJIWARA
;    APPLICANT:  Takeshi, WATANABE
;    APPLICANT:  Masato, HORIE
;    APPLICANT:  Toyomasa, KATAGIRI
;    TITLE OF INVENTION:  HUMAN GENE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/820,170A
;      FILING DATE:
;      CLASSIFICATION:  536
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1493 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA(genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    IMMEDIATE SOURCE:
;      LIBRARY:  Human fetal brain cDNA library
;      CLONE:  GEN-025F07
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  95..478
US-08-820-170A-9


  Query Match             52.0%;  Score 13;  DB 2;  Length 1493;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgttt 13
           |||||||||||||
Db     443 GGATCTGCTGTTT 431


RESULT  42
US-09-055-699-9/c
; Sequence 9, Application US/09055699
; Patent No. 6005088
;  GENERAL INFORMATION:
;    APPLICANT:  Tsutomu, FUJIWARA
;    APPLICANT:  Takeshi, WATANABE
;    APPLICANT:  Masato, HORIE
;    APPLICANT:  Toyomasa, KATAGIRI
;    TITLE OF INVENTION:  HUMAN GENE
;    NUMBER OF SEQUENCES:  42
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
```

```
;       STREET:  2100 Pennsylvania Avenue, N.W.
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  United States
;       ZIP:  20037-3202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/055,699
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/820,170
;       FILING DATE:
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 293-7060
;       TELEFAX:  (202) 293-7860
;       TELEX:  6491103
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1493 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA(genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     IMMEDIATE SOURCE:
;       LIBRARY:  Human fetal brain cDNA library
;       CLONE:  GEN-025F07
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  95..478
US-09-055-699-9


  Query Match             52.0%;  Score 13;  DB 3;  Length 1493;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgttt 13
          |||||||||||||
Db    443 GGATCTGCTGTTT 431


RESULT  43
US-09-273-565-9/c
; Sequence 9, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
;  APPLICANT: FUJIWARA, TSUTOMU
;  APPLICANT: WATANABE, TAKESHI
;  APPLICANT: HORIE, MASATO
;  TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
;  TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
;  FILE REFERENCE: Q-53599
;  CURRENT APPLICATION NUMBER: US/09/273,565A
;  CURRENT FILING DATE: 1999-03-22
;  EARLIER APPLICATION NUMBER: 09/055,699
;  EARLIER FILING DATE: 1998-04-07
;  EARLIER APPLICATION NUMBER: 08/820,170
;  EARLIER FILING DATE: 1997-03-19
;  EARLIER APPLICATION NUMBER: JP 63410/1996
;  EARLIER FILING DATE: 1996-03-19
;  EARLIER APPLICATION NUMBER: JP 69163/1997
;  EARLIER FILING DATE: 1997-03-05
;  NUMBER OF SEQ ID NOS: 95
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
;   LENGTH: 1493
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (95)..(478)
US-09-273-565-9


  Query Match             52.0%;  Score 13;  DB 4;  Length 1493;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgttt 13
          |||||||||||||
Db    443 GGATCTGCTGTTT 431


RESULT  44
US-08-171-299B-1
; Sequence 1, Application US/08171299B
; Patent No. 5599665
;  GENERAL INFORMATION:
;     APPLICANT:  Barbieri, Joseph T.
;     APPLICANT:  Frank, Dara W.
;     TITLE OF INVENTION:  EXOENZYME S PROTEIN PREPARATION AND CLONED    EXOENZYME S
;     NUMBER OF SEQUENCES:  19
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  QUARLES & BRADY
;       STREET:  411 East Wisconsin Avenue
;       CITY:  Milwaukee
;       STATE:  Wisconsin
;       COUNTRY:  U.S.A.
;       ZIP:  53202-4497
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/171,299B
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Baker, Jean C.
;       REGISTRATION NUMBER:  35,433
;       REFERENCE/DOCKET NUMBER:  650053.90871
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (414) 277-5709
;       TELEFAX:  (414) 271-3552
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1712 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-171-299B-1


  Query Match             52.0%;  Score 13;  DB 1;  Length 1712;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 tgaagcgcagcag 25
           |||||||||||||
Db    388 TGAAGCGCAGCAG 400


RESULT  45
US-09-221-235-9/c
; Sequence 9, Application US/09221235
; Patent No. 6043040
; GENERAL INFORMATION:
```

```
;  APPLICANT: Acton, Susan
;  TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-050
;  CURRENT APPLICATION NUMBER: US/09/221,235
;  CURRENT FILING DATE: 1998-12-28
;  EARLIER APPLICATION NUMBER: 09/163,115
;  EARLIER FILING DATE:
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 1743
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1743)
US-09-221-235-9


  Query Match           52.0%;  Score 13;  DB 3;  Length 1743;
  Best Local Similarity  100.0%;  Pred. No. 62;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 tgaagcgcagcag 25
           |||||||||||||
Db      64 TGAAGCGCAGCAG 52


Search completed: December 26, 2001, 12:52:37
Job time: 6789 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2001, 14:10:04 ; Search time 4619.78 Seconds
(without alignments)
58.151 Million cell updates/sec

Title: US-09-396-196F-4
Perfect score: 25
Sequence: 1 ggatctgctgtttgaagcgcagcag 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 9

Total number of hits satisfying chosen parameters: 752014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	17	68.0	381	11	BG888762	BG888762 EST514613
c	2	17	68.0	479	10	BE508108	BE508108 dc26f09.y
	3	16	64.0	293	13	AZ214122	AZ214122 Sheared D
	4	16	64.0	358	10	AI207085	AI207085 qf59a09.x
c	5	16	64.0	360	11	C61070	C61070 C61070 Yuji
	6	16	64.0	502	10	BE515828	BE515828 WHE0606_B
	7	16	64.0	517	13	AQ692054	AQ692054 HS_5408_A
c	8	16	64.0	555	13	AQ889100	AQ889100 HS_2204_A
	9	16	64.0	558	10	AW064984	AW064984 ST38C02 P
c	10	16	64.0	567	10	BE232412	BE232412 137424 MA
	11	16	64.0	632	13	AZ449344	AZ449344 1M0247J04
c	12	16	64.0	650	13	TA69C05P	AL457556 T. brucei
	13	15	60.0	177	13	FR0044714	AL132206 Fugu rubr
	14	15	60.0	190	11	BI024308	BI024308 CM3-MT029
c	15	15	60.0	278	10	AI021028	AI021028 ua99b01.r
	16	15	60.0	298	10	BB337949	BB337949 BB337949
c	17	15	60.0	316	13	FR0039786	AL127281 Fugu rubr
c	18	15	60.0	358	11	BF355075	BF355075 RC2-HT082
	19	15	60.0	372	13	AQ521765	AQ521765 HS_5239_A
c	20	15	60.0	406	13	AQ143500	AQ143500 HS_3066_B
	21	15	60.0	424	13	AQ494880	AQ494880 HS_5195_A
	22	15	60.0	427	13	BH077581	BH077581 RPCI-24-2
	23	15	60.0	435	13	AQ114851	AQ114851 CIT-HSP-2
	24	15	60.0	437	10	AW432846	AW432846 sh81g05.y
c	25	15	60.0	456	11	BF555582	BF555582 UI-R-E0-c
c	26	15	60.0	466	10	AI477772	AI477772 fb54h11.y
	27	15	60.0	479	13	AZ965014	AZ965014 2M0234K24
	28	15	60.0	495	13	AQ192174	AQ192174 HS_3227_B
c	29	15	60.0	507	10	BE114310	BE114310 UI-R-BJ1-
c	30	15	60.0	509	10	AW915326	AW915326 EST346630
c	31	15	60.0	528	10	AA818607	AA818607 UI-R-A0-b
c	32	15	60.0	532	13	TA375E11P	AL495591 T. brucei
c	33	15	60.0	542	10	AA790463	AA790463 vw04g12.r
c	34	15	60.0	559	13	TA162C10P	AL472440 T. brucei
	35	15	60.0	594	13	FR0036852	AL124355 Fugu rubr
	36	15	60.0	598	13	FR0035666	AL123183 Fugu rubr
	37	15	60.0	600	13	FR0044579	AL132071 Fugu rubr
c	38	15	60.0	611	13	FR0044632	AL132124 Fugu rubr
c	39	15	60.0	621	11	BG306774	BG306774 fm06h08.y
c	40	15	60.0	627	11	BG017095	BG017095 df68g03.x
	41	15	60.0	632	13	AQ754510	AQ754510 HS_5434_A
	42	15	60.0	640	11	BG546580	BG546580 602574435
c	43	15	60.0	644	11	BG969048	BG969048 602835003
	44	15	60.0	645	10	BE597936	BE597936 PI1_67_C0
c	45	15	60.0	649	11	BG085090	BG085090 H3108E05-
	46	15	60.0	663	10	AW633493	AW633493 b108d10.w
c	47	15	60.0	681	11	BE820949	BE820949 GM700013A
	48	15	60.0	716	11	BG971333	BG971333 602839869
c	49	15	60.0	748	10	AW128741	AW128741 fe36h03.y
c	50	15	60.0	752	13	CNS02O6Q	AL206315 Tetraodon
c	51	15	60.0	846	11	BG428569	BG428569 602501093
c	52	15	60.0	870	13	CNS04GGY	AL289627 Tetraodon
	53	15	60.0	878	11	BG721419	BG721419 602694974
	54	15	60.0	888	13	CNS04KH2	AL294815 Tetraodon
c	55	15	60.0	893	13	CNS04YX0	AL313533 Tetraodon
c	56	15	60.0	922	13	CNS02FQO	AL195369 Tetraodon
c	57	15	60.0	961	13	CNS01I50	AL145285 Anopheles
c	58	15	60.0	966	13	CNS01T1Q	AL165959 Tetraodon
c	59	15	60.0	977	13	CNS05364	AL319045 Tetraodon
c	60	15	60.0	1044	13	CNS05B8I	AL329499 Tetraodon
	61	14	56.0	138	11	BG189673	BG189673 RST8719 A
	62	14	56.0	139	13	AZ613372	AZ613372 1M0441M11
	63	14	56.0	144	11	BF955843	BF955843 PM2-NN120
c	64	14	56.0	161	11	BI389308	BI389308 949052E12
	65	14	56.0	205	10	AA884873	AA884873 am21c11.s
c	66	14	56.0	211	10	BE456581	BE456581 ut11c03.x
	67	14	56.0	216	10	BB200491	BB200491 BB200491
c	68	14	56.0	227	10	BE012248	BE012248 945016A03
	69	14	56.0	229	11	BF369300	BF369300 CM4-GN008
	70	14	56.0	230	13	AQ506129	AQ506129 RPCI-11-3
c	71	14	56.0	236	11	D24951	D24951 RICR2836A R
	72	14	56.0	241	10	AV040895	AV040895 AV040895
	73	14	56.0	247	11	BG217212	BG217212 RST36914
	74	14	56.0	251	10	BB601253	BB601253 BB601253
	75	14	56.0	261	10	AV360182	AV360182 AV360182
c	76	14	56.0	268	10	AV002603	AV002603 AV002603
	77	14	56.0	271	11	BG308817	BG308817 fl89d01.x
c	78	14	56.0	281	10	AI882699	AI882699 fb15c08.y
	79	14	56.0	282	11	BG182292	BG182292 RST1157 A
	80	14	56.0	285	13	AZ084513	AZ084513 RPCI-23-3
c	81	14	56.0	288	10	BB125447	BB125447 BB125447
c	82	14	56.0	300	11	N47165	N47165 yy73c07.r1
	83	14	56.0	303	10	BB276182	BB276182 BB276182
c	84	14	56.0	305	10	AV136350	AV136350 AV136350
c	85	14	56.0	307	11	BG439729	BG439729 dab26d02.

86	14	56.0	312	10	AA623286	AA623286 vi19f05.r
87	14	56.0	313	11	BI188823	BI188823 d3d06fs.r
88	14	56.0	316	10	BB154393	BB154393 BB154393
89	14	56.0	322	10	BB338422	BB338422 BB338422
90	14	56.0	332	11	BG198869	BG198869 RST18139
91	14	56.0	334	11	BG193289	BG193289 RST12417
92	14	56.0	337	10	AW127031	AW127031 ga20b07.y
93	14	56.0	343	10	AW245381	AW245381 2822875.3
94	14	56.0	345	10	AI708068	AI708068 as60f01.x
c 95	14	56.0	345	10	BB165235	BB165235 BB165235
96	14	56.0	347	10	AI302370	AI302370 qn49g09.x
97	14	56.0	348	10	AA902490	AA902490 ok91g12.s
98	14	56.0	362	10	AI669961	AI669961 wc11b06.x
99	14	56.0	362	10	AI810191	AI810191 wf80h12.x
c 100	14	56.0	363	10	AA418282	AA418282 zv96g01.s

ALIGNMENTS

RESULT 1
BG888762/c
LOCUS BG888762 381 bp mRNA EST 30-MAY-2001
DEFINITION EST514613 cSTD Solanum tuberosum cDNA clone cSTD11O14 5' sequence, mRNA sequence.
ACCESSION BG888762
VERSION BG888762.1 GI:14265848
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 381)
AUTHORS van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chieminqo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE Generations of ESTs from dormant potato tubers
JOURNAL Unpublished (2001)
COMMENT Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.
FEATURES Location/Qualifiers
source 1. .381
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTD11O14"
/clone_lib="cSTD"
/tissue_type="dormant tuber"
/dev_stage="one month post-harvest"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; This library targets genes expressed in dormant tubers. This library was made from sections of dormant tuber, avoiding the buds and epidermis. Tubers were stored for one month post-harvest at 4oC. The tuber was peeled, well away from the surface. Then it was chopped into 1-2 mm cubes and immediately frozen in liquid nitrogen. This library is noted as P4 in Tanksley lab notebooks."
BASE COUNT 88 a 76 c 111 g 106 t
ORIGIN

Query Match 68.0%; Score 17; DB 11; Length 381;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       9 tgtttgaagcgcagcag 25
           |||||||||||||||||
Db     311 TGTTTGAAGCGCAGCAG 295
```

RESULT 2
BE508108/c
LOCUS BE508108 479 bp mRNA EST 07-AUG-2000
DEFINITION dc26f09.y1 NICHD XGC Li1 Xenopus laevis cDNA clone IMAGE:3398249 5' similar to TR:O75940 O75940 30KDA SPLICING FACTOR. ;, mRNA sequence.
ACCESSION BE508108
VERSION BE508108.1 GI:9727883
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 479)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 434.
FEATURES Location/Qualifiers
source 1. .479
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3398249"
/clone_lib="NICHD XGC Li1"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.4 kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library."
BASE COUNT 114 a 147 c 112 g 106 t
ORIGIN

Query Match 68.0%; Score 17; DB 10; Length 479;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       9 tgtttgaagcgcagcag 25
           |||||||||||||||||
Db     183 TGTTTGAAGCGCAGCAG 167
```

RESULT 3
AZ214122
LOCUS AZ214122 293 bp DNA GSS 09-JUN-2000
DEFINITION Sheared DNA-67G1.TF Sheared DNA Trypanosoma brucei genomic clone Sheared DNA-67G1, DNA sequence.
ACCESSION AZ214122
VERSION AZ214122.1 GI:8431922
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 293)
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library

```
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: Sheared DNA-67G1.TR
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
            DNA library constructed at TIGR. Clones will be available for
            distribution through Research Genetics, Alabama, USA. Sheared DNA
            end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/.
            Seq primer: M13-Forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .293
                     /organism="Trypanosoma brucei"
                     /strain="TREU927/4 GUTat 10.1"
                     /db_xref="taxon:5691"
                     /clone="Sheared DNA-67G1"
                     /clone_lib="Sheared DNA"
                     /note="Vector: pUC18; Site_1: SmaI; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from a cloned population of
                     Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
                     sheared to give a tight size distribution (approx 2 kb).
                     The v + i method used for the library construction is
                     described in detail in Smith, H.O. and Venter, J.C.
                     (Making small insert libraries for whole genome shotgun
                     sequencing projects.  In Genome Sequencing: A Practical
                     Approach, eds. M. Vaudin and B. Barell, Oxford University
                     Press, 1999)."
BASE COUNT       38 a     73 c     78 g    104 t
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 293;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 atctgctgtttgaagc 18
           ||||||||||||||||
Db     149 ATCTGCTGTTTGAAGC 164


RESULT   4
AI207085
LOCUS       AI207085      358 bp    mRNA            EST       10-NOV-1998
DEFINITION  qf59a09.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1754296
            3' similar to TR:Q61655 Q61655 EUKARYOTIC TRANSLATION INITIATION
            FACTOR 4A RELATED SEQUENCE 1 ;, mRNA sequence.
ACCESSION   AI207085
VERSION     AI207085.1  GI:3765757
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 358)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 652   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 239.
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1754296"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']."
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       98 a     79 c     73 g    107 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 358;
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       8 ctgtttgaagcgcagc 23
           ||||||||||||||||
Db     294 CTGTTTGAAGCGCAGC 309


RESULT   5
C61070/c
LOCUS       C61070        360 bp    mRNA            EST       22-SEP-1997
DEFINITION  C61070 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk224c7 5', mRNA sequence.
ACCESSION   C61070
VERSION     C61070.1  GI:2419775
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
            ,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="taxon:6239"
                     /clone="yk224c7"
                     /clone_lib="Yuji Kohara unpublished cDNA"
                     /sex="hermaphrodite, male"
                     /tissue_type="whole animal"
                     /dev_stage="varied"
BASE COUNT      101 a     71 c     76 g    111 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 11;  Length 360;
```

```
  Best Local Similarity   100.0%;  Pred. No. 68;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 atctgctgtttgaagc 18
          ||||||||||||||||
Db     26 ATCTGCTGTTTGAAGC 11


RESULT   6
BE515828
LOCUS       BE515828      502 bp    mRNA            EST       08-AUG-2000
DEFINITION  WHE0606_B07_C14ZA Wheat ABA-treated embryo cDNA library Triticum
            aestivum cDNA clone WHE0606_B07_C14, mRNA sequence.
ACCESSION   BE515828
VERSION     BE515828.1  GI:9739842
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 502)
  AUTHORS   Anderson,O.D., Chao,S., Han,P.S., Hsia,C.C., Johnson,R.R, Kang,Y.,
            Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Tong,J.C., Verhey
            ,S.D. and Walker-Simmons,M.K.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - ABA-treated embryo library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Clontech Matchmaker 3' AD primer.
FEATURES             Location/Qualifiers
     source          1. .502
                     /organism="Triticum aestivum"
                     /cultivar="Brevor (soft, white, winter, common wheat)"
                     /db_xref="taxon:4565"
                     /clone="WHE0606_B07_C14"
                     /clone_lib="Wheat ABA-treated embryo cDNA library"
                     /tissue_type="Seed embryo"
                     /dev_stage="Mature dormant seeds"
                     /lab_host="E. coli DH12S"
                     /note="Vector: pGAD10; Site_1: EcoRI; Site_2: XhoI;
                     Embryos were cut from mature, dormant seeds and imbibed in
                     25 microM ABA  (abscisic acid) in 5 mM Mes buffer, pH 5.7,
                     for 12 hr at 22 C.  The  tissue, total RNA, and poly(A)
                     RNA were prepared by Steven Verhey in  M.K.
                     Walker-Simmons's lab (USDA-ARS, Washington State Univ.,
                     Pullman,  Washington 99164-6420.  A cDNA library was made
                     by Clontech using a  combination of random and oligo dT
                     primers. Library was plated and  archived by Russell
                     Johnson (Colby College, ME/Walker-Simmons' lab).  Plasmid
                     DNA preparations and DNA sequencing were performed in the
                     OD  Anderson lab (all other authors)."
BASE COUNT      103 a    108 c    128 g    162 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 502;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 atctgctgtttgaagc 18
          ||||||||||||||||
Db    189 ATCTGCTGTTTGAAGC 204


RESULT   7
AQ692054
LOCUS       AQ692054      517 bp    DNA             GSS       06-JUL-1999
DEFINITION  HS_5408_A2_E07_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=984 Col=14 Row=I, DNA sequence.
ACCESSION   AQ692054
VERSION     AQ692054.1  GI:5382302
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 984  row: I  column: 14
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 517.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=984 Col=14 Row=I"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      154 a     96 c     83 g    178 t      6 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 517;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 atctgctgtttgaagc 18
          ||||||||||||||||
Db    411 ATCTGCTGTTTGAAGC 426


RESULT   8
AQ889100/c
LOCUS       AQ889100      555 bp    DNA             GSS       10-NOV-1999
DEFINITION  HS_2204_A1_C07_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2204 Col=13 Row=E, DNA sequence.
ACCESSION   AQ889100
VERSION     AQ889100.1  GI:6345290
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2204  row: E  column: 13
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 555.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2204 Col=13 Row=E"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      131 a    135 c    147 g    124 t     18 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 13;  Length 555;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaagcg 19
          ||||||||||||||||
Db    497 TCTGCTGTTTGAAGCG 482


RESULT   9
AW064984
LOCUS       AW064984      558 bp    mRNA            EST       12-OCT-1999
DEFINITION  ST38C02 Pine TriplEx shoot tip library Pinus taeda cDNA clone
            ST38C02, mRNA sequence.
ACCESSION   AW064984
VERSION     AW064984.1  GI:6020056
KEYWORDS    EST.
SOURCE      loblolly pine.
  ORGANISM  Pinus taeda
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
  TITLE     The Pine Gene Discovery Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ross Whetten
            Forest Biotechnology Group
            North Carolina State University
            Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
            ,NC, 27695-8008
            Tel: 919-515-7800
            Fax: 919-515-7801
            Email: rosswhet@unity.ncsu.edu
            Seq primer: 5' lambda TriplEx2 Sequencing Primer.
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Pinus taeda"
                     /db_xref="taxon:3352"
                     /clone="ST38C02"
                     /clone_lib="Pine TriplEx shoot tip library"
                     /lab_host="E. coli BM25.8"
                     /note="Organ: shoot tips; Vector: Lambda TriplEx; Site_1:
                     SfiI (A); Site_2: SfiI (B); Shoot tips (approx. 2 cm from
                     apex) were collected during the spring, frozen and used
                     for mRNA isolation. The SMART-PCR method (Clontech) was
                     used to prepare a library from 1 ug total RNA, using the
                     Lambda TriplEx vector. Plasmid subclones in pTriplEx were
                     recovered by cre-lox excision in E. coli strain BM25.8 and
                     sequenced from the 5' end."
BASE COUNT      146 a    117 c    131 g    160 t      4 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 558;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 gtttgaagcgcagcag 25
           ||||||||||||||||
Db     108 GTTTGAAGCGCAGCAG 123


RESULT  10
BE232412/c
LOCUS       BE232412      567 bp    mRNA            EST       10-JUL-2000
DEFINITION  137424 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BE232412
VERSION     BE232412.1  GI:9017130
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 72  row: E  column: 24
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT      114 a    198 c    171 g     83 t      1 others
ORIGIN


  Query Match             64.0%;  Score 16;  DB 10;  Length 567;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       8 ctgtttgaagcgcagc 23
           ||||||||||||||||
Db     227 CTGTTTGAAGCGCAGC 212
```

```
RESULT  11
AZ449344
LOCUS       AZ449344      632 bp    DNA             GSS       04-OCT-2000
DEFINITION  1M0247J04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0247J04 R, DNA sequence.
ACCESSION   AZ449344
VERSION     AZ449344.1  GI:10603038
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0247  row: J  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 632.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0247J04"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      185 a    167 c    110 g    170 t
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 13;  Length 632;
  Best Local Similarity   100.0%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttgaa 16
           ||||||||||||||||
Db     552 GGATCTGCTGTTTGAA 567
```

```
RESULT  12
TA69C05P/c
LOCUS       TA69C05P      650 bp    DNA             GSS       13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 69c05, forward sequence,
            genomic survey sequence.
ACCESSION   AL457556
VERSION     AL457556.1  GI:11858782
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 650)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="69c05"
BASE COUNT      208 a    186 c    173 g     83 t
ORIGIN
```

```
  Query Match             64.0%;  Score 16;  DB 13;  Length 650;
  Best Local Similarity   100.0%;  Pred. No. 71;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 atctgctgtttgaagc 18
           ||||||||||||||||
Db     357 ATCTGCTGTTTGAAGC 342
```

```
RESULT  13
FR0044714
LOCUS       FR0044714     177 bp    DNA             GSS       22-OCT-1999
DEFINITION  Fugu rubripes GSS sequence, clone 192G14eE6, genomic survey
            sequence.
ACCESSION   AL132206
VERSION     AL132206.1  GI:6114152
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
```

```
REFERENCE   1  (bases 1 to 177)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .177
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 192G14"
                     /clone="192G14eE6"
BASE COUNT       33 a     41 c     50 g     44 t      9 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 177;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 atctgctgtttgaag 17
           ||||||||||||||||
Db     116 ATCTGCTGTTTGAAG 130


RESULT  14
BI024308
LOCUS       BI024308      190 bp    mRNA            EST       14-JUN-2001
DEFINITION  CM3-MT0293-260101-692-f03 MT0293 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI024308
VERSION     BI024308.1  GI:14430938
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-MT0293-
            260101-692-f03&t3=2001-01-26&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 80
            High quality sequence stop: 175.
FEATURES             Location/Qualifiers
     source          1. .190
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MT0293"
                     /dev_stage="Adult"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       22 a     49 c     64 g     54 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 190;
  Best Local Similarity   100.0%;  Pred. No. 2.3e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ctgctgtttgaagcg 19
           |||||||||||||||
Db      44 CTGCTGTTTGAAGCG 58


RESULT  15
AI021028/c
LOCUS       AI021028      278 bp    mRNA            EST       16-JUN-1998
DEFINITION  ua99b01.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1365577 5' similar to TR:O35626 O35626 RAS-RELATED PROTEIN.
            ;, mRNA sequence.
ACCESSION   AI021028
VERSION     AI021028.1  GI:3235364
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 278)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:898797
            Possible reversed clone: similarity on wrong strand
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 106.
FEATURES             Location/Qualifiers
     source          1. .278
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1365577"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
```

adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 72 a 74 c 82 g 50 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 278;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatctgctgtttga 15
||||||||||||||
Db 18 GGATCTGCTGTTTGA 4

RESULT 16
BB337949
LOCUS BB337949 298 bp mRNA EST 11-JUL-2000
DEFINITION BB337949 RIKEN full-length enriched, 10 days neonate cerebellum Mus musculus cDNA clone B930010K12 3', mRNA sequence.
ACCESSION BB337949
VERSION BB337949.1 GI:9046712
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 298)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M. , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y. , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y. , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I. , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino ,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .298
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="B930010K12"
/clone_lib="RIKEN full-length enriched, 10 days neonate cerebellum"
/sex="mixed"
/tissue_type="cerebellum"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
BASE COUNT 81 a 52 c 56 g 109 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 298;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 atctgctgtttgaag 17
|||||||||||||||
Db 12 ATCTGCTGTTTGAAG 26

RESULT 17
FR0039786/c
LOCUS FR0039786 316 bp DNA GSS 22-OCT-1999
DEFINITION Fugu rubripes GSS sequence, clone 103O02bB10, genomic survey sequence.
ACCESSION AL127281
VERSION AL127281.1 GI:6108896
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 316)
AUTHORS Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K., Umrania,Y., Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB. UK Email: biohelp@hgmp.mrc.ac.uk
COMMENT Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic sequence.
FEATURES Location/Qualifiers
source 1. .316
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 103O02"
/clone="103O02bB10"
BASE COUNT 85 a 86 c 79 g 65 t 1 others

```
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 316;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 atctgctgtttgaag 17
          ||||||||||||||||
Db    149 ATCTGCTGTTTGAAG 135


RESULT  18
BF355075/c
LOCUS       BF355075      358 bp    mRNA            EST       22-NOV-2000
DEFINITION  RC2-HT0827-240700-014-h01 HT0827 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF355075
VERSION     BF355075.1  GI:11314149
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 358)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC2&t2=RC2-HT0827-
            240700-014-h01&t3=2000-07-24&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 13
            High quality sequence stop: 358.
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="HT0827"
                     /dev_stage="Adult"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       88 a    111 c     81 g     78 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 358;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaagc 18
          ||||||||||||||||
Db    275 TCTGCTGTTTGAAGC 261


RESULT  19
AQ521765
LOCUS       AQ521765      372 bp    DNA             GSS       05-MAY-1999
DEFINITION  HS_5239_A2_H09_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=815 Col=18 Row=O, DNA sequence.
ACCESSION   AQ521765
VERSION     AQ521765.1  GI:4752890
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 372)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 815  row: O  column: 18
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 372.
FEATURES             Location/Qualifiers
     source          1. .372
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=815 Col=18 Row=O"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       97 a     86 c     72 g    115 t      2 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 372;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 gatctgctgtttgaa 16
          ||||||||||||||||
Db    315 GATCTGCTGTTTGAA 329


RESULT  20
AQ143500/c
LOCUS       AQ143500      406 bp    DNA             GSS       24-SEP-1998
DEFINITION  HS_3066_B2_H10_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3066 Col=20 Row=P, DNA sequence.
ACCESSION   AQ143500
VERSION     AQ143500.1  GI:3534153
KEYWORDS    GSS.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3066  row: P  column: 20
            Class: BAC ends
            High quality sequence stop: 406.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3066 Col=20 Row=P"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      138 a      78 c      54 g     135 t       1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 406;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 gatctgctgtttgaa 16
           ||||||||||||||||
Db      98 GATCTGCTGTTTGAA 84


RESULT  21
AQ494880
LOCUS       AQ494880      424 bp    DNA             GSS       28-APR-1999
DEFINITION  HS_5195_A1_B06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=771 Col=11 Row=C, DNA sequence.
ACCESSION   AQ494880
VERSION     AQ494880.1  GI:4695047
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 424)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 771  row: C  column: 11
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 424.
FEATURES             Location/Qualifiers
     source          1. .424
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=771 Col=11 Row=C"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      137 a      70 c      82 g     134 t       1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 424;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaagc 18
           |||||||||||||||
Db      91 TCTGCTGTTTGAAGC 105


RESULT  22
BH077581
LOCUS       BH077581      427 bp    DNA             GSS       18-JUL-2001
DEFINITION  RPCI-24-291N2.TJ RPCI-24 Mus musculus genomic clone RPCI-24-291N2,
            DNA sequence.
ACCESSION   BH077581
VERSION     BH077581.1  GI:14897178
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-291N2.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 291  row: N  column: 2
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-291N2"
```

```
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      102 a    106 c     93 g    126 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 427;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 tctgctgtttgaagc 18
          |||||||||||||||
Db    326 TCTGCTGTTTGAAGC 340


RESULT  23
AQ114851
LOCUS       AQ114851      435 bp    DNA             GSS       29-AUG-1998
DEFINITION  CIT-HSP-2376L11.TF CIT-HSP Homo sapiens genomic clone 2376L11, DNA
            sequence.
ACCESSION   AQ114851
VERSION     AQ114851.1  GI:3490972
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 435)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2376L11.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .435
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2376L11"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT       87 a    115 c    118 g    115 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 435;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 gatctgctgtttgaa 16
          |||||||||||||||
Db    375 GATCTGCTGTTTGAA 389


RESULT  24
AW432846
LOCUS       AW432846      437 bp    mRNA            EST       11-FEB-2000
DEFINITION  sh81g05.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1016-6609 5' similar to TR:Q40171 Q40171 TDR6 MRNA ;, mRNA
            sequence.
ACCESSION   AW432846
VERSION     AW432846.1  GI:6964153
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 350.
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-6609"
                     /clone_lib="Gm-c1016"
                     /tissue_type="immature flowers of field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from immature flowers of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT      133 a    101 c     83 g    120 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 437;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 gatctgctgtttgaa 16
          |||||||||||||||
```

```
Db    123 GATCTGCTGTTTGAA 137


RESULT  25
BF555582/c
LOCUS       BF555582      456 bp    mRNA            EST       12-DEC-2000
DEFINITION  UI-R-E0-cy-d-07-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone
            UI-R-E0-cy-d-07-0-UI 5', mRNA sequence.
ACCESSION   BF555582
VERSION     BF555582.1  GI:11665312
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 456)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1770501
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1. .456
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-E0-cy-d-07-0-UI"
                     /clone_lib="UI-R-E0"
                     /dev_stage="embryonic"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: NotI; Site_2: EcoRI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from 8, 12 and 18-day embryo.  The
                     tag is a string  of 3-5 nucleotides present between the
                     Not I site and the oligo-dT track  which allows
                     identification of the library of origin of a clone within
                     the mixture."
BASE COUNT      112 a    115 c    139 g     88 t      2 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 456;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttga 15
          ||||||||||||||||
Db     88 GGATCTGCTGTTTGA 74


RESULT  26
AI477772/c
LOCUS       AI477772      466 bp    mRNA            EST       07-JUN-2001
DEFINITION  fb54h11.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone
            IMAGE:3715749 5', mRNA sequence.
ACCESSION   AI477772
VERSION     AI477772.1  GI:4335383
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 465
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:3715749"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT3']
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT       97 a    135 c    120 g    113 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 466;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 ctgtttgaagcgcag 22
          ||||||||||||||||
Db    330 CTGTTTGAAGCGCAG 316


RESULT  27
AZ965014
```

```
LOCUS       AZ965014      479 bp    DNA             GSS       27-APR-2001
DEFINITION  2M0234K24R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0234K24 R, DNA sequence.
ACCESSION   AZ965014
VERSION     AZ965014.1  GI:13836241
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0234  row: K  column: 24
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 479.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0234K24"
                     /clone_lib="Mouse 10kb plasmid UUGC2M library"
                     /sex="Female"
                     /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (female) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      163 a     91 c    122 g    103 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 479;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 gatctgctgtttgaa 16
           ||||||||||||||||
Db     197 GATCTGCTGTTTGAA 211


RESULT  28
AQ192174
LOCUS       AQ192174      495 bp    DNA             GSS       04-NOV-1998
DEFINITION  HS_3227_B1_G01_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3227 Col=1 Row=N, DNA sequence.
ACCESSION   AQ192174
VERSION     AQ192174.1  GI:3590796
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3227  row: N  column: 1
            Class: BAC ends
            High quality sequence stop: 495.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3227 Col=1 Row=N"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      135 a    114 c     79 g    167 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 495;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 gatctgctgtttgaa 16
           ||||||||||||||||
Db     316 GATCTGCTGTTTGAA 330


RESULT  29
BE114310/c
LOCUS       BE114310      507 bp    mRNA            EST       13-JUN-2000
DEFINITION  UI-R-BJ1-awk-g-02-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
            UI-R-BJ1-awk-g-02-0-UI 3', mRNA sequence.
ACCESSION   BE114310
VERSION     BE114310.1  GI:8506415
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
```

```
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized heart library cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: clones will be available through Research
            Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES              Location/Qualifiers
     source           1. .507
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-BJ1-awk-g-02-0-UI"
                      /clone_lib="UI-R-BJ1"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
                      library is a subtracted library derived from the following
                      tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
                      canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
                      AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
                      For a detailed description of the library from which this
                      clone was derived, please visit our web site at
                      ratest.eng.uiowa.edu. The subtraction has been previously
                      described in (Bonaldo, Lennon and Soares, Genome Research
                      6:791-806, 1996)
                      TAG_LIB=UI-R-BJ1
                      TAG_TISSUE=heart
                      TAG_SEQ=ACAAC"
BASE COUNT      125 a    119 c    105 g    158 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 507;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 tctgctgtttgaagc 18
           ||||||||||||||||
Db     357 TCTGCTGTTTGAAGC 343


RESULT  30
AW915326/c
LOCUS       AW915326      509 bp    mRNA            EST       25-MAY-2000
DEFINITION  EST346630 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
            RGICU11 5' end, mRNA sequence.
ACCESSION   AW915326
VERSION     AW915326.1  GI:8081017
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            This clone is available through the ATCC, contact the ATCC
            tel#703-365-2700 for further information
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .509
                      /organism="Rattus sp."
                      /db_xref="taxon:10118"
                      /clone="RGICU11"
                      /clone_lib="Normalized rat embryo, Bento Soares"
                      /dev_stage="embryo 8, 12, 18 dpc"
                      /note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      116 a    167 c    118 g    107 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 10;  Length 509;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ggatctgctgtttga 15
           ||||||||||||||||
Db     443 GGATCTGCTGTTTGA 429


RESULT  31
AA818607/c
LOCUS       AA818607      528 bp    mRNA            EST       03-JUL-1999
DEFINITION  UI-R-A0-bc-h-05-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
            UI-R-A0-bc-h-05-0-UI 3' similar to gb|U96683|MMU96683 Mus musculus
            immunoglobulin-like receptor PIRA2 (6M23) mRNA, partial cds, mRNA
            sequence.
ACCESSION   AA818607
VERSION     AA818607.1  GI:4228398
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     On Feb 17, 1998 this sequence version replaced gi:2889346.
            Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence tag present in the cDNA between the NotI site and the
            oligo-dT track served to identify it as a clone from the normalized
            adult Kidney library. cDNA Library Preparation: M. Fatima Bonaldo,
            Ph.D. Clone distribution: clones will be available through Research
            Genetics This clone is also available through the I.M.A.G.E.
            Consortium at LLNL (info@image.llnl.gov). IMAGE ID=1768272
            Seq primer: M13 Forward
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .528
                      /organism="Rattus norvegicus"
                      /strain="Sprague-Dawley"
                      /db_xref="taxon:10116"
                      /clone="UI-R-A0-bc-h-05-0-UI"
                      /clone_lib="UI-R-A0"
                      /dev_stage="adult"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
```

polylinker; Site_1: Not I; Site_2: Eco RI; This library consists of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain , liver, kidney, heart, spleen, ovary, and muscle. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."
BASE COUNT 138 a 131 c 101 g 158 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 528;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tctgctgtttgaagc 18
||||||||||||||||
Db 353 TCTGCTGTTTGAAGC 339

RESULT 32
TA375E11P/c
LOCUS TA375E11P 532 bp DNA GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 375e11, forward sequence, genomic survey sequence.
ACCESSION AL495591
VERSION AL495591.1 GI:11873033
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
REFERENCE 1 (bases 1 to 532)
AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh1@sanger.ac.uk
COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (
4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES Location/Qualifiers
source 1. .532
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="375e11"
BASE COUNT 162 a 156 c 92 g 122 t
ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 532;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ctgctgtttgaagcg 19
|||||||||||||||
Db 310 CTGCTGTTTGAAGCG 296

RESULT 33
AA790463/c
LOCUS AA790463 542 bp mRNA EST 06-FEB-1998
DEFINITION vw04g12.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1230982 5' similar to SW:RAP3_HUMAN P17964 RAS-RELATED PROTEIN RAP-2B. ;, mRNA sequence.
ACCESSION AA790463
VERSION AA790463.1 GI:2850583
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 542)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:656574
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 492.
FEATURES Location/Qualifiers
source 1. .542
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1230982"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 127 a 178 c 125 g 112 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 542;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatctgctgtttga 15
|||||||||||||||
Db 486 GGATCTGCTGTTTGA 472

RESULT 34
TA162C10P/c
LOCUS TA162C10P 559 bp DNA GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 162c10, forward sequence, genomic survey sequence.
ACCESSION AL472440

```
VERSION     AL472440.1  GI:11837892
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nhl@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="162c10"
BASE COUNT      198 a    154 c     92 g    115 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 559;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ctgctgtttgaagcg 19
           |||||||||||||||
Db     188 CTGCTGTTTGAAGCG 174


RESULT  35
FR0036852
LOCUS       FR0036852     594 bp    DNA             GSS       22-OCT-1999
DEFINITION  Fugu rubripes GSS sequence, clone 039N10aE9, genomic survey
            sequence.
ACCESSION   AL124355
VERSION     AL124355.1  GI:6105970
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 039N10"
                     /clone="039N10aE9"
BASE COUNT      138 a    135 c    144 g    157 t     20 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 594;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 atctgctgtttgaag 17
           |||||||||||||||
Db     475 ATCTGCTGTTTGAAG 489


RESULT  36
FR0035666
LOCUS       FR0035666     598 bp    DNA             GSS       22-OCT-1999
DEFINITION  Fugu rubripes GSS sequence, clone 019B02cC2, genomic survey
            sequence.
ACCESSION   AL123183
VERSION     AL123183.1  GI:6104798
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 019B02"
                     /clone="019B02cC2"
BASE COUNT      120 a    147 c    153 g    165 t     13 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 598;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 atctgctgtttgaag 17
           |||||||||||||||
Db     116 ATCTGCTGTTTGAAG 130


RESULT  37
FR0044579
LOCUS       FR0044579     600 bp    DNA             GSS       22-OCT-1999
DEFINITION  Fugu rubripes GSS sequence, clone 192G14aF9, genomic survey
            sequence.
ACCESSION   AL132071
VERSION     AL132071.1  GI:6114017
KEYWORDS    GSS; genome survey sequence.
```

```
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 192G14"
                     /clone="192G14aF9"
BASE COUNT      123 a    152 c    153 g    170 t      2 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 600;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 atctgctgtttgaag 17
          ||||||||||||||||
Db    145 ATCTGCTGTTTGAAG 159


RESULT  38
FR0044632/c
LOCUS       FR0044632     611 bp    DNA             GSS       22-OCT-1999
DEFINITION  Fugu rubripes GSS sequence, clone 192G14bC7, genomic survey
            sequence.
ACCESSION   AL132124
VERSION     AL132124.1  GI:6114070
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 192G14"
                     /clone="192G14bC7"
BASE COUNT      158 a    154 c    150 g    122 t      27 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 611;
  Best Local Similarity   100.0%;  Pred. No. 2.4e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 atctgctgtttgaag 17
          ||||||||||||||||
Db    450 ATCTGCTGTTTGAAG 436


RESULT  39
BG306774/c
LOCUS       BG306774      621 bp    mRNA            EST       22-FEB-2001
DEFINITION  fm06h08.y1 Zebrafish adult retina cDNA Danio rerio cDNA clone
            4145535 5' similar to TR:Q9Z0X1 Q9Z0X1 APOPTOSIS-INDUCING FACTOR
            AIF. ;, mRNA sequence.
ACCESSION   BG306774
VERSION     BG306774.1  GI:13104301
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Rasborinae; Danio.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            Library constructed by: Susan E. Brockerhoff DNA Sequencing by:
            Washington University Genome Sequencing Center Clone distribution:
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 420.
FEATURES             Location/Qualifiers
     source          1. .621
                     /organism="Danio rerio"
                     /strain="wild-type"
                     /db_xref="taxon:7955"
                     /clone="4145535"
                     /clone_lib="Zebrafish adult retina cDNA"
                     /sex="mixed"
                     /dev_stage="1-2 years"
                     /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"
                     /note="Vector: Lambda ZAP II (pBluescript SK-); Site_1:
                     EcoRI; Site_2: SalI; This Zebrafish library was
                     constructed by Dr. Susan E. Brockerhoff (email:
                     sbrocker@u.washington.edu) RZPD library number: 760"
BASE COUNT      157 a    141 c    164 g    154 t      5 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 621;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 ctgtttgaagcgcag 22
          ||||||||||||||||
Db    180 CTGTTTGAAGCGCAG 166
```

```
RESULT  40
BG017095/c
LOCUS       BG017095      627 bp    mRNA            EST       18-APR-2001
DEFINITION  df68g03.x1 Xenopus laevis unfertilized egg cDNA library Xenopus
            laevis cDNA clone IMAGE:3744580 3', mRNA sequence.
ACCESSION   BG017095
VERSION     BG017095.1  GI:12470841
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 627)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: df68g03.y1
            Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Bruce Blumberg
             Library normalized by Jihwan Song
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clone distribution information for
            this library can be found through Research Genetics, visit their
            web page at: http://www.resgen.com/ Please reference the id listed
            below when ordering this clone: Source lab clone id - xlnneg005n05
            Seq primer: -40UP from Gibco
            High quality sequence stop: 468.
FEATURES             Location/Qualifiers
     source          1. .627
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:3744580"
                     /clone_lib="Xenopus laevis unfertilized egg cDNA library"
                     /tissue_type="unfertilized egg"
                     /lab_host="Top-10 F'"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from 2ug of poly A+ RNA.
                     EcoRI-XhoI cut cDNA was then ligated into UniZap-XR
                     (Stratagene) with EcoRI at the 5' end and XhoI at the 3'
                     end. SS-library phagemids were prepared by mass excision
                     from the original library and normalized by  hybridization
                     to biotinylated driver (prepared from the same library by
                     PCR) to Cot-omega of 11. After removal of hybrids and
                     excess driver by streptavidin sepharose chromatography,
                     the  ss-phagemids were made double stranded and
                     electroporated into Top-10 F'. Original library
                     contruction by Bruce Blumberg (Blumberg et al., 1991
                     Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
                     2923-2935). Note: This is a Xenopus Gene Collection (XGC)
                     library."
BASE COUNT      146 a    131 c    143 g    206 t      1 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 627;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 tttgaagcgcagcag 25
           |||||||||||||||
Db     556 TTTGAAGCGCAGCAG 542
```

```
RESULT  41
AQ754510
LOCUS       AQ754510      632 bp    DNA             GSS       27-JUL-1999
DEFINITION  HS_5434_A1_G05_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=1010 Col=9 Row=M, DNA sequence.
ACCESSION   AQ754510
VERSION     AQ754510.1  GI:5618726
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 1010  row: M  column: 9
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 632.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=1010 Col=9 Row=M"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      156 a    143 c    126 g    190 t     17 others
ORIGIN


  Query Match             60.0%;  Score 15;  DB 13;  Length 632;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       7 gctgtttgaagcgca 21
           |||||||||||||||
Db     308 GCTGTTTGAAGCGCA 322


RESULT  42
BG546580
LOCUS       BG546580      640 bp    mRNA            EST       04-APR-2001
DEFINITION  602574435F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4702473 5',
            mRNA sequence.
ACCESSION   BG546580
VERSION     BG546580.1  GI:13545245
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1539  row: f  column: 10
            High quality sequence stop: 640.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4702473"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      170 a    131 c    155 g    184 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 atctgctgtttgaag 17
          |||||||||||||||
Db    600 ATCTGCTGTTTGAAG 614


RESULT  43
BG969048/c
LOCUS       BG969048      644 bp    mRNA            EST       12-JUN-2001
DEFINITION  602835003F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4989312 5',
            mRNA sequence.
ACCESSION   BG969048
VERSION     BG969048.1  GI:14356685
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 644)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11003  row: f  column: 01
            High quality sequence stop: 642.
FEATURES             Location/Qualifiers
     source          1. .644
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4989312"
                     /clone_lib="NCI_CGAP_Co24"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      153 a    196 c    164 g    131 t
ORIGIN


  Query Match             60.0%;  Score 15;  DB 11;  Length 644;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ggatctgctgtttga 15
          |||||||||||||||
Db    470 GGATCTGCTGTTTGA 456


RESULT  44
BE597936
LOCUS       BE597936      645 bp    mRNA            EST       18-AUG-2000
DEFINITION  PI1_67_C08.g1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
            mRNA sequence.
ACCESSION   BE597936
VERSION     BE597936.1  GI:9853009
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: pathogen-induced plants
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 3
            High quality sequence stop: 635
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .645
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Pathogen induced 1 (PI1)"
                     /note="Organ: Anthracnose-infected leaves from
                     two-week-old sorghum plants 48 hr after inoculation;
                     Vector: pBluescript II from Lambda Zap II; Site_1: XhoI;
                     Site_2: EcoRI; Two-week-old sorghum plants (BTX 623
                     cultivar) were infected with pathogen (isolate FRM42I of
                     Colletotrichum graminicola, which is a sorghum isolate).
                     RNA was prepared from infected leaves harvested from 45
                     seedlings  48 hours after inoculation.  Note: young
                     seedlings (2 weeks old) exhibit juvenile resistant
                     reaction, which is an incompatible interaction. As they
                     grow older (4 weeks or older), plants resume susceptiblity
```

to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
BASE COUNT 179 a 143 c 170 g 153 t
ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 645;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 ctgctgtttgaagcg 19
||||||||||||||||
Db 627 CTGCTGTTTGAAGCG 641

RESULT 45
BG085090/c
LOCUS BG085090 649 bp mRNA EST 26-JAN-2001
DEFINITION H3108E05-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone H3108E05 5', mRNA sequence.
ACCESSION BG085090
VERSION BG085090.1 GI:12567654
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 649)
AUTHORS Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka,T.S., Carter,M.G. and Ko,M.S.H.
TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: H3108E05-3
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3108 row: E column: 05
Seq primer: -21M13 Reverse
High quality sequence stop: 649
POLYA=No.
FEATURES Location/Qualifiers
source 1. .649
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="H3108E05"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse develolpment, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
BASE COUNT 157 a 182 c 177 g 133 t
ORIGIN

Query Match 60.0%; Score 15; DB 11; Length 649;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatctgctgtttga 15
||||||||||||||||
Db 315 GGATCTGCTGTTTGA 301

Search completed: December 26, 2001, 14:10:10
Job time: 7057 sec